AUGUST
1954
VOLUME 18
NUMBER 4

JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF CONSULTING PSYCHOLOGY

Edited by Laurance F. Shaffer

Teachers College, Columbia University

Associate Editors: Edgar A. Doll, *Bellingham, Wash.* Ann Magaret Garner, *University of Illinois.* William A. Hunt, *Northwestern University.* E. Lowell Kelly, *University of Michigan,* Morris Krugman, *Board of Education, New York, N. Y.* Bertha M. Luckey, *Board of Education, Cleveland, Ohio.* Fred McKinney, *University of Missouri.* Anne Roe, *New York, N. Y.* Carl R. Rogers, *University of Chicago.* R. Nevitt Sanford, *University of California.* Edward Joseph Shoben, Jr., *Teachers College, Columbia University,* J. R. Wittenborn, *Yale University.*

Lorraine Bouthilet, *Managing Editor*

The *Journal of Consulting Psychology* is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problem of clinical psychology also appear in its page.

Editorial Office

Manuscripts submitted for publication, books and tests for review, and correspondence on editorial matters should be sent to the Editor, Laurance F. Shaffer, 525 West 120th St., New York 27, N. Y.

Information for contributors. The important characteristics of a good article are the significance of its subject matter, the demonstrated soundness of its conclusions, its clarity, and its brevity. Clarity is improved by careful outlining, discriminating use of headings, and a simple, straightforward style of writing. For brevity, authors should strive to convey every thought without waste of words, and should avoid repetitions of methods, data, or findings.

The style of all manuscripts should conform to the instructions given in the Publication Manual of the American Psychological Association, prepared by the Council of Editors (*Psychological Bulletin,* 1952, 49, 389-449). If it is not otherwise available, the *Manual* may be purchased, for $1.00, from the Washington office of the Association. The Editor expects each author to study the *Manual* carefully before submitting a manuscript. Special attention should be given to the style of headings, capitalization, indention, spacing, and footnotes, and to the preparation of tables, figures, and lists of references. Otherwise acceptable manuscripts which depart seriously from the standards of the *Manual* are returned to the author for revision and retyping.

There is a lag of about nine to twelve months between the acceptance of an article and its publication. Authors may secure immediate publication, however, by arranging for articles to be printed as extra pages, increasing the number of pages that subscribers receive. The cost to the author of such early publication is approximately $20.00 per printed page.

Authors receive fifty reprints gratis, and may order additional quantities when returning proof.

Business Office

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters, should be sent to the American Psychological Association, Inc., 1333 Sixteenth St. N. W., Washington 6, D. C. Address changes must reach the subscription office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

The *Journal of Consulting Psychology* is published bimonthly in February, April, June, August, October, and December, the yearly volume comprising approximately 480 pages. The subscription per year is $7.00, foreign $7.50, single number $1.50.

Copyright, 1954 by the American Psychological Association, Inc.

Printed and mailed by the Dentan Printing Co., 23 West Colorado Ave., Colorado Springs, Colo. and entered as second-class matter at the Post Office at Colorado Springs, Colo. Acceptance for mailing at special rate of postage provided for in section 538, Act of February 25, 1925, authorized May 27, 1940.

JOURNAL OF CONSULTING PSYCHOLOGY

August, 1954 Vol. 18, No. 4

CONTENTS

The Clinician as Judge: Implications of a Series of Studies: *Roy M. Hamlin* - - - - 233

The Clinician as Judge: Details of Procedure in Judging Projective Material: *Irving Bialick and Roy M. Hamlin* - - - - - - - - - 239

The Clinician as Judge: Judgments of Adjustment from Rorschach Single-Card Performance: *S. Thomas Cummings* - - - - - - - - - - - - 243

The Clinician as Judge: Total Rorschachs and Clinical Case Material: *Richard L. Newton* - - - - - - - - - - - - - 248

Intra-Individual Consistency in Response to Tests of Varying Degrees of Ambiguity: *Arthur C. Carr* - - - - - - - - - - - - - - - - - - - 251

Rorschach Concepts of Normality: *Ann Lawler Brockway, Goldine C. Gleser, and George A. Ulett* - - - - - - - - - - - - - - - 259

An Evaluation of the Memory-for-Designs Test: *Alvin R. Howard and Donald J. Shoemaker* - - - - - - - - - - - - - - 266

Determinant Shift of Psychoneurotic and Psychotic Patients: *Robert G. Gibby and Bernard A. Stotsky* - - - - - - - - - - - - - - - 267

Rorschach Card Rejection in Psychodiagnosis: *Ivan N. Mensh and Joseph D. Matarazzo* 271

Intelligence Test Reliability for Cerebral Palsied Children: *David H. Crowell and Doris C. Crowell* - - - - - - - - - - - - - - - 276

A Cross Validation of Davidson's Rorschach Adjustment Scale: *Raymond J. Corsini and Harold F. Uehling* - 277

Failure to Return for Treatment in Tested and Nontested Clinic Patients: *Miriam McCue, Morris Goodman, and Melvin Rosenthal* - - - - - - - - - - 280

An Evaluation of Psychological Reports in a Clinical Setting: *Sol L. Garfield, Ralph W. Heine, and Morton Leventhal* - - - - - - - - - 281

Some Common Misconceptions About Neuroticism: *Harrison G. Gough* - - - - - - 287

Factors Descriptive of Chronic Schizophrenics Selected for the Operation of Prefrontal Lobotomy: *Maurice Lorr, Richard L. Jenkins, and James Q. Holsopple* - 293

Group Therapy Techniques with Hospitalized Chronic Schizophrenic Patients: *Joseph M. Sacks and Stanley Berger* - - - - - - - - - - - - - - - - - - - 297

The Influence of Fullness of Interview on the Reliability, Discriminability, and Validity of Interview Judgments: *Harry Levin* - - - - - - - - - - - - - - - - - 303

New Books and Tests - 307

Brief Reports:
A New Publication Procedure

For a trial period, the *Journal of Consulting Psychology* will accept *"brief reports"* for early publication without expense to the author, under certain conditions. The new procedure is intended to reduce publication lag and permit the circulation of worth-while studies of specialized interest or minor importance which cannot be accepted at present because of limitations of space.

A *brief report,* strictly limited to one printed page, may be accepted if the author agrees to make a more extended report available. Up to four pages in each issue may be devoted to brief reports, published in the order of their receipt without regard to the dates of receipt of articles accepted for regular publication. It is anticipated that most of the brief reports can appear in the first or second issue to go to press following their acceptance.

The procedure for submitting a *brief report* is:

1. The author sends the editor *both* the brief report and a more complete report of the study. The full report should be of sufficient length to give a clear account of the background, procedure, results, and conclusions of the study.

2. The *full report* is prepared in the style specified by the *Publication Manual* [1], except that it may be typed with single spacing. The editor will send the full report to the ADI.

3. The *brief report* gives a clear, condensed summary of the procedure of the study, and as full an account of the results as space permits. To insure that the brief report does not exceed one printed page, the author prepares it according to these specifications:

The text of the brief report, including all matter except the title and the author's lines, does not exceed 80 typewritten lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, or 4.2 inches long in pica. The manuscript is double spaced throughout and, except for the short lines, follows the standard style [1].

Headings, tables, and references are avoided in the brief report. If essential, they must be counted in the 80 lines. The brief report includes, within its 80 lines, a footnote in this style:[1]

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. [giving the author's full name and address] or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. .... from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $..... for microfilm or $..... for photocopies.

4. The author of a brief report prepares at least 100 mimeographed copies of the extended report. He agrees: (*a*) to send copies of the extended report upon request as long as the supply lasts, and (*b*) not to submit the extended report for publication to another printed journal of general circulation.

Reference

1. American Psychological Association. Council of Editors. Publication manual of the American Psychological Association. *Psychol. Bull.*, 1952, 49, 389-449.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

The Clinician as Judge: Implications of a Series of Studies

Roy M. Hamlin
Western Psychiatric Institute, University of Pittsburgh

In their most useful application, projective techniques are not concentrated estimates of single variables. They are not even multiscore tests. They permit extremely complex, unstructured behavior. The number and selection of variables are so little controlled that the behavior cannot conceivably be represented by measuring each ultimate factor separately. Statisticians like Cronbach [6] have attempted to suggest the mathematical complexity involved. Langmuir [16], the physicist, has stated that we must face the concept of "divergent phenomena," events so complexly determined that we cannot study them at all within the framework of classical science. He feels we must recognize the limits of science, and instead of science use common sense, judgment, and intuition.

At the present time, clinical psychologists use judgment. They describe and predict events or tendencies in the field of personality. For this purpose, they may use projective techniques as complex and particularly meaningful samples of individual behavior. The true value of these procedures is not represented in a single score or in a series of scores.

The reliability and validity of such techniques are not readily demonstrated. However, if psychologists can be persuaded to make limited judgments based on these techniques, the judgments can be tested for reliability and for validity. To a degree we can check the usefulness and accuracy of the procedures involved. In doing so, we follow the common procedure of science: from something extremely complex we tease out something simple enough to deal with objectively.

The studies of judgment with projective material have, so far, been somewhat disappointing. The reports on projective techniques in Kelly and Fiske's Michigan study of embryo psychologists [14], the Air Force studies using the Rorschach [20], the generally excellent California Institute of Child Welfare study of judgments of adjustment in a group of adolescent subjects [8], Kurtz's carefully fair use of the Rorschach in attempting to select insurance company personnel [15], and other similar studies have left the impression with both opponents and friends of projective techniques that such studies have so far shown zero or close to zero correlation between judgments based on projective techniques and any acceptably established outside criterion.

Purpose

The present discussion is limited chiefly to 10 studies, which used either so-called Rorschach signs, or judgments based on projective and clinical material. Five of these studies yielded positive results, and five failed to do so. The positive results were not entirely discouraging, and the negative results may be due to specific details of method. A consideration of these studies suggests that many factors contribute to positive or negative results. Chiefly for purposes of emphasis, however, one factor may be singled out for comment and stated in the form of a hypothesis.

Hypothesis: Positive and negative results in these studies are directly related to the simplicity or complexity of the material judged, and to the adequacy of the experimental conditions in enabling the clinician to derive judgments from the degree of complexity presented. ***Subhypothesis 1***: In general, negative results may be expected when the units of material judged are too simple or atomistic. ***Subhypothesis 2***:

In general, negative results may be expected when the units of material judged are too complex to be manageable within the conditions of the experiment.

It is probably apparent that the general thesis offered here is in some opposition, at least, to the use of so-called Rorschach signs and, on the other hand, to uncritical emphasis of global units or macrounits in clinical research. This opposition may be more apparent than real.

The Studies

The 10 studies considered were carried out by staff and graduate students associated with the University of Pittsburgh and with the Veterans Administration in the Pittsburgh area. Of these studies, four used Rorschach signs and six used judgments by clinicians. The studies have been reported elsewhere [1, 2, 4, 5, 7, 10, 11, 13, 18, 19, 21], and the two least available [13, 19] have been summarized by Newton [18]. Three other references to the series [3, 9, 12] are listed at the end of this article. Briefly, the 10 studies may be identified as follows:

1. Hamlin, Albee, and Leland [10] compared Muench's Rorschach signs for 71 subjects, divided into three groups of socially adjusted subjects, maladjusted subjects, and neuropsychiatric patients. The sign scores failed to distinguish between the groups.

2. Hamlin, Berger, and Cummings [11] compared Muench's Rorschach signs for 20 patients before and after six months or more of psychiatric therapy. Results were essentially negative though minimally in keeping with expectation.

3. Basic Rorschach Scores on the same 20 subjects [11] showed an average gain in adjustment score following therapy of 8 points, significant at the .01 level of confidence.

4. Newton and Goodman [19] compared Basic Rorschach Scores on 50 subjects with two different criteria of adjustment. Results were not significant.

5. Albee and Hamlin [1, 2] used a paired-comparison method to develop a scale of adjustment judged from drawings of a man and woman. Using this scale with 72 cases, judges rated normal subjects as significantly better adjusted than neurotics, and as significantly better adjusted than schizophrenics. From these data, Bendig has derived a correlation of .47, significant at the .01 level of confidence between judgments and criterion groups.

6. Bialick [4, 5] used a paired-comparison method to develop a similar scale of intelligence judged only from five Rorschach *W*, or whole, responses for each subject. In details of method, this study is the most interesting of the series. For a small sample of 25 subjects, judgments and Wechsler-Bellevue IQ correlated .68, significant at the .01 level of confidence.

7. Cummings [7] developed 10 scales of adjustment, one for each Rorschach card. With 50 cases, the 10 combined judgments on each subject correlated .50 with an outside criterion of adjustment.

8. Newton [18] developed a scale of adjustment judged from total Rorschachs. With 50 cases, the Rorschach judgments failed to show a significant correlation with an outside criterion identical with Cummings', and with another outside criterion.

9. Hamlin and Newton [13] presented two Rorschachs from Newton's study to a seminar for "blind" analysis. In Newton's previous study, the two Rorschachs had been rated at exactly the same level of good adjustment, although one subject was known to be socially well adjusted and the other was psychotic and markedly maladjusted in nearly every respect. No other basis was considered in selecting the subjects. After careful study of the Rorschachs only, each of the eight members of the seminar rated as less well adjusted the subject known to be so, and the average group rating of the degree of difference in adjustment on Newton's seven-point scale was "reasonably accurate."

10. Whitmyre [21] had psychologists judge adjustment from drawings, and art students judge the same drawings for artistic quality. Each judge sorted and ranked 50 drawings as a single task, and an illustrated judgment scale was not used. The implications of the resulting high correlation (.78) between adjustment values and art values will not be considered in this paper. However, an incidental validity correlation between psychologists' judgments of cases and an outside criterion of adjustment was essentially negative.

Complexity of Units

The main thesis of this paper is related to the complexity of units judged. A rough continuum may be set up for the simplicity or complexity of the material:

1. Rorschach signs or other scored or judged single units that represent only small fragments of the projective material, mechanically selected.

2. Figure drawings or parts of the Rorschach selected on the basis of some rationale and retaining some degree of organization or configurational meaningfulness, some degree of complexity.

3. Total Rorschachs or units of similar complexity.

These three levels of units are illustrated by the studies of Newton [18], Cummings [7], Newton and Goodman [19], and Hamlin and Newton [13]. Newton used total Rorschachs, Cummings used single Rorschach cards, Newton and Goodman used Rorschach signs, and Hamlin and Newton presented two of Newton's 50 cases for blind analysis to a seminar where the protocols could be studied intensively and carefully. These four studies were essentially comparable in the judgment required, in the outside criterion, in the kind of subjects used, and in the projective technique considered. The four studies differed in the simplicity and complexity of the units judged, and in the comparable conditions provided for making the judgments.

Using Rorschach signs applied to Newton's 50 cases, Newton and Goodman found no significant relation with the relatively adequate outside criterion. The signs selected for this study were Buhler's Basic Rorschach Score.

With 10 different scales, one for each Rorschach card, and 10 separate judgments, one for each card, Cummings' resulting correlation of .50 was significant at the .01 level of confidence. The units judged were not too complex for the conditions of judgment.

However, when Newton used total Rorschachs as the basis of judgment, with entirely comparable subjects, judges, and outside criterion, no significant validity correlation was found. This result seems contrary to the idea that global judgments are more meaningful than judgments based on somewhat smaller units. However, the contradiction is not necessarily a real one. The one important factor that may be stressed is that the judging procedure broke down. The seven-point judgment scale was illustrated by a Rorschach at each point on the scale. These sample illustrations were too complex to be useful. No one judge rated more than 15 Rorschachs, and the judges spent ample time and effort on these. However, increasing complexity beyond a certain point probably does not lead to meaningful global understanding, except under ideal conditions or in a clinical situation where data can be cross-checked, discarded as not pertinent, or synthesized in a manner that eventually results in a simplified picture derived from complexity.

Stated in a crude way, the judgment process should be about as complex as the material judged and as the judgment required. For example, our seminar study of two of Newton's cases at least suggests that complex units of projective material, in this case total Rorschachs, can be used successfully in making judgments. These two Rorschachs had been rated at exactly the same level of adjustment by Newton's judges. However, under seminar conditions, every member of the seminar rated as better adjusted the subject so identified by the outside criterion, diagnosis, and over-all clinical picture.

In general, the results of all 10 studies may be interpreted as supporting the main thesis: Small, mechanically scored or judged units tend to give negative results. The likelihood of positive results is increased by larger units, by samples of behavior that are sufficient to represent some degree of organization, so that a clinical judgment—rather than a mechanical scoring without a rationale—is possible. However, the global unit, or macrounit, will give negative results unless the conditions for judging are complex, time-consuming, and approach actual clinical procedures. Briefly, the studies may be grouped in terms of unit size and judgment conditions:

Atomistic units. Of the four studies with Rorschach signs [10, 11, 19], three gave negative results. The one statistically significant result was probably due to an unusually good outside criterion.

Simple configurations. All three studies in which a judgment scale was established first and then applied to units smaller than global Rorschachs [1, 2, 4, 5, 7] gave results significant at the .01 level of confidence. Judgments of drawings correlated .47 with adjustment; judgments of single Rorschach cards correlated .50 with adjustment; and judgments of five *W* responses correlated .68 with intelligence.

Complex units. The three remaining studies [13, 18, 21] suggest the precautions that must be taken if the judge is presented with extensive, complex material in a single judgment task. Judgments of total Rorschachs gave negative validity results. Ranking 50 drawings as a single task, and without an illustrated judgment scale, gave negative validity results. Time-consuming, intensive analysis of two complete Rorschachs did permit correct judgments, but in these two cases only and under conditions that are impractical for many research purposes.

The California Child Welfare study of Rorschach judgments and adjustment [8] closely parallels one of our studies [18] in procedure and negative results. Other studies that have given negative results, and received wide attention, have used total Rorschachs or equally complex projective material, without providing ideal conditions for analyzing, digesting, and converting this material into the judgments required.

Other Considerations

Obviously, the complexity of units judged is not the only factor in determining positive or negative results. It is represented in the 10 studies considered, and it raises a more general question: Do we expect too much from clinical "experts"? A few other factors may be mentioned as of particular interest.

Degree of relationship expected. In planning a study in which clinicians' judgments of projective material are to be correlated against some second criterion, we should make a realistic estimate in advance of the degree of correlation that may reasonably be expected. Many psychologists tend to take a naive attitude toward clinicians and toward projective material. Clinicians *say* they can judge adjustment or introversive tendency from the Rorschach. Therefore, we expect their judgments to correlate at least .70 with any haphazard outside criterion: psychiatric diagnosis as a measure of adjustment, or a paper-and-pencil test of something that happens to be labelled introversion. We expect clinical "experts" to look at Rorschachs, for a relatively brief time, with material often dubiously reproduced, and then to pick good clinical psychologists with no idea what a good clinical psychologist is, or insurance sales managers with no rationale and no real clinical experience with these individuals. We tend to expect magic.

In an area where methodology is little developed, Cummings' correlation of .50 with a dubious outside criterion and an entity like adjustment, which has hardly been defined, is surprisingly high. Using only five Rorschach responses per subject, Bialick obtained a correlation of .68 with Wechsler IQ. Judgments on a mere fragment of the Rorschach correlate better with the Wechsler than does the CAVD, the Otis, or several other respectable tests of intelligence.

Definition and outside criterion. In regard to defining a variable and in regard to selecting and establishing an outside criterion, we should consider the known skills and limitations of the clinician. Bialick used intelligence as a variable. The judges were experienced clinicians with extensive knowledge of the meaning and measurement of intelligence. They were specifically acquainted with the operations involved in the test used as an outside criterion, the Wechsler. Furthermore, the outside criterion was a carefully constructed test of intelligence. It was not a sample of complex, real-life behavior only partially reflecting intelligence.

In using the variable "adjustment," we were on less firm ground. Adjustment was defined verbally and in terms of an outside criterion. The verbal definitions were sketchy, but in the later studies they tended more and more to include such phrases as "freedom from anxiety," "ability to handle anxiety," "contact with reality," "capacity for delusions or hallucinations." That is, we simply used other words to parallel our outside criterion groups: normals, neurotics, and psychotics. It might have been better to describe in precise and repetitive detail what the outside criterion was. The outside criterion closely paralleled the extensive and recent experience of the

judges. The judges were already acquainted with the Rorschach performances of many neurotic and psychotic patients.

Material judged. Complexity of units has been discussed. One other point might be stressed in regard to material judged. With projective techniques especially, an adequate sample of material pertinent to the judgment required should be assured.

Bialick's study may be used for illustration. In the Rorschach, the blots and procedure do not require the subject to give responses within the framework of a specific variable to be judged. For this reason, a subject may not produce a sample of behavior pertinent to a specific judgment study. Bialick studied the quality of *W* responses as related to intelligence. Therefore, he took every Rorschach that had five or more *W* responses, but discarded any Rorschach that had fewer *W* responses. McReynolds [17] studied form level, and to do so had his subjects pass on responses that he selected in advance. The point is, for many research purposes, controls should be used to obtain an adequate sample of pertinent behavior. A well-integrated paranoid schizophrenic can give an innocuous Rorschach by default, and for clinical purposes his ability to do so is important. For some research purposes, however, we might do well to remember Rorschach's own comments on the need for control studies and control procedures. Ratios, correlational corrections, or "cannot say" categories for the judges are no substitute for an adequate sample of behavior.

Judges. In our expectations and plans for a judgment study, the clinician judge should be regarded as a subject, not a magician. His assets and limitations should be considered in decisions on complexity of task, sample of behavior presented, definition of variable, and selection of outside criterion. In addition to his background of general experience, he should have specific training in definition of variable, and should be given a trial run with correction of errors and opportunity to review his mistakes. Eventually, the basis of successful judgments should be thoroughly studied.

We might assume that the "expertness" of a competent clinician consists of a high degree of skill in descriptive prediction for subjects and variables with which he is extensively and recently acquainted in a somewhat limited setting. A good state hospital psychologist can probably say much that is valid and useful about state hospital patients. Put him in a correctional institution or a clinic for children and his predictive validity may be limited for a time. Use him as a judge in one of our studies and his predictive validity may be limited indeed. His skills may be readily converted or transplanted, but this process must be fostered and provided for. In this paper, complexity of units may illustrate the necessity for studying the clinician carefully in setting the judgment task.

Conclusions

In summary, the present paper has offered several suggestions in regard to studies of clinical judgment applied to projective techniques:

1. Such judgments are not like traditional test scores. In traditional tests, complex material is reduced to mechanical units that are counted. With projective techniques, the clinician may be asked to make an arbitrarily simplified judgment. Both methods serve to tease out something simple from something complex. But the methods are not the same.

2. A number of widely publicized studies have produced negative results in applying judgments to projective material. Usually, overly complex units of projective material have been employed.

3. In a few studies we have obtained positive results by using samples of projective behavior that were neither overly simple nor overly complex.

4. The positive results obtained in two of these studies have represented encouraging degrees of correlation. At the present stage, Cummings' correlation of .50 with a crude criterion of adjustment, and Bialick's correlation of .68 with Wechsler IQ may be considered respectable.

5. Incidental comment has been made on other methodological considerations. All comments emphasize careful study of the clinician in setting the judgment task.

Received August 17, 1953.

References

1. Albee, G. W., & Hamlin, R. M. An investigation of the reliability and validity of judgments

of adjustment inferred from drawings. *J. clin. Psychol.*, 1949, **5**, 396-398.

2. Albee, G. W., & Hamlin, R. M. Judgment of adjustment from drawings: the applicability of rating scale methods. *J. clin. Psychol.*, 1950, **6**, 363-365.
3. Albee, G. W., & Hamlin, R. M. The place of judgment in clinical research. *J. clin. Psychopathol.*, 1950, **11**, 174-177.
4. Bialick, I. A method of applying clinical judgment to problems of Rorschach validation: the quality of Rorschach whole responses and intelligence. Unpublished master's thesis, Univer. of Pittsburgh, 1950.
5. Bialick, I., & Hamlin, R. M. The clinician as judge: details of procedure in judging projective material. *J. consult. Psychol.*, 1954, **18**, 239-242.
6. Cronbach, L. J. Statistical methods for multiscore tests. *J. clin. Psychol.*, 1950, **6**, 21-25.
7. Cummings, S. T. The clinician as judge: judgments of adjustment inferred from Rorschach single-card performance. *J. consult. Psychol.*, 1954, **18**, 243-247.
8. Grant, Marguerite Q., Ives, Virginia, & Ranzoni, Jane H. Reliability and validity of judges' ratings of adjustment on the Rorschach. *Psychol. Monogr.*, 1952, **66**, No. 2. (Whole No. 334).
9. Hamlin, R. M., & Albee, G. W. Muench's tests before and after nondirective therapy: a control group for his subjects. *J. consult. Psychol.*, 1948, **12**, 412-416.
10. Hamlin, R. M., Albee, G. W., & Leland, E. M. Objective Rorschach "signs" for groups of normal, maladjusted and neuropsychiatric subjects. *J. consult. Psychol.*, 1950, **14**, 276-282.
11. Hamlin, R. M., Berger, B., & Cummings, S. T. Changes in adjustment following psychotherapy as reflected in Rorschach signs. In W. Wolff, & J. A. Precker (Eds.), *Success in psychotherapy*. New York: Grune & Stratton, 1952. Pp. 94-111.
12. Hamlin, R. M., & Kogan, W. S. Objectification in Rorschach interpretation. *J. Pers.*, 1948, **17**, 177-181.
13. Hamlin, R. M., & Newton, R. L. Comparison of a schizophrenic and a normal subject, both rated by clinicians as well-adjusted on the basis of "blind" Rorschachs. Paper read at Eastern Psychol. Ass., Atlantic City, N. J., March, 1952.
14. Kelly, E. L., & Fiske, D. W. The prediction of success in the VA training program in clinical psychology. *Amer. Psychologist*, 1950, **5**, 395-406.
15. Kurtz, A. K. A research test of the Rorschach test. *Personnel Psychol.*, 1948, **1**, 41-51.
16. Langmuir, I. Science, common sense, and decency. *Science*, 1943, **97**, 1-7.
17. McReynolds, P. Perception of Rorschach concepts as related to personality deviations. *J. abnorm. soc. Psychol.*, 1951, **46**, 131-141.
18. Newton, R. L. The clinician as judge: total Rorschachs and clinical case material. *J. consult. Psychol.*, 1954, **18**, 248-250.
19. Newton, R. L., & Goodman, H. W. A comparison of Basic Rorschach Scores with judgments of adjustment based on clinical case material. Paper read at Eastern Psychol. Ass., Atlantic City, N. J., March, 1952.
20. Sells, S. B. Description of the USAF School of Aviation Medicine research program on the development of psychiatric screening devices for selection of flying personnel. Paper read at Amer. Psychol. Ass., Washington, September, 1952.
21. Whitmyre, J. W. The significance of artistic excellence in the judgment of adjustment inferred from human figure drawings. *J. consult. Psychol.*, 1953, **17**, 421-424.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

The Clinician as Judge: Details of Procedure in Judging Projective Material [1, 2]

Irving Bialick
VA Hospital, Nashville, Tenn.
and Roy M. Hamlin
Western Psychiatric Institute, University of Pittsburgh

This paper reports an objective study indicating the correlation between judgments of intelligence based on Rorschach whole, or *W*, responses and Wechsler-Bellevue IQ. However, the study was planned essentially to illustrate and consider certain details of procedure in the use of judgment scales with projective material. In the section titled *Details of Method* we will highlight and comment briefly on specific considerations.

The hypothesis tested in the study was: experienced clinical psychologists can make valid judgments of intelligence based on five Rorschach *W* responses for each subject.

Subjects and Judges

The subjects of the study proper were 25 outpatients of the Pittsburgh VA Mental Hygiene Clinic. All were male, white veterans of World War II. The mean chronological age was 28 years, with a range from 20 to 39 years. All subjects had a psychiatric diagnosis of neurosis, with no reason to suspect the presence of schizophrenic reaction or of organic intracranial involvement. The Rorschach and Wechsler were administered as part of the routine psychological evaluation. No subject was included who had not produced at least five *W* responses on the Rorschach. The subjects were further selected to provide a distribution of IQ's exactly proportionate to Wechsler's intelligence classification of the general population. No other basis of selection was used.

In addition to these 25 subjects, 10 other subjects were used in the construction of a judgment scale by the method of paired comparisons. Except in distribution of IQ, these 10 subjects were comparable in all respects to the 25 subjects. In terms of IQ, the 10 subjects were selected so that two subjects had Wechsler IQ's within three points of each of the following IQ levels: 80, 100, 115, 125, and 135.

The judges in the study were four VA staff psychologists, thoroughly trained and experienced with both Rorschach and Wechsler. For purposes of incidental comparison, four less experienced VA psychologists in training were used as a second group of judges. None of the judges had any prior knowledge of the cases used in the study.

The Judgment Scale

In the construction of the judgment scale, 10 subjects, 2 at each of five IQ levels, were used. From the Rorschach of each subject, five *W* responses were selected according to an arbitrary system. These responses, including inquiry, were typed on 5 × 8 cards. The 10 cards were then presented to four experienced judges in the 45 possible pairings. The pairs were presented in randomized order to two judges, and in reverse order to the other two judges. Each judge considered two cards at a time and decided which set of *W* responses had been given by the more intelligent subject.

From this paired-comparison method, the 10 subjects were arranged in order of intelli-

[1] This article is based in part on a master's thesis submitted by Dr. Bialick in partial fulfillment of the requirements for the M.S. degree, University of Pittsburgh, 1950.

[2] Published with the permission of the Chief Medical Director, Department of Medicine and Surgery, Veterans Administration, who assumes no responsibility for the opinions expressed or the conclusions drawn by the authors.

gence according to each judge's estimates. Rho correlations of validity and reliability, for these 10 subjects only, will be reported here rather than under *Results*. The rhos for individual judges against intelligence ranged from .71 to .87. The combined rankings of all four judges yielded a rho of .77, significant at the .01 level of confidence. The reliability rho for the ratings of two judges against two judges was .87, significant at the .01 level. These correlations were based on 10 cases, with two cases selected arbitrarily at each of five IQ levels.

Only 5 of the 10 subjects were used in illustrating the final judgment scale. At each of the five known IQ levels, that case was retained which had been judged correctly by the clinicians, and which best represented a distinct step interval. The five cases retained were ordered by the judges in complete agreement with the criterion of Wechsler IQ. Finally, a numerical scale value was calculated for each scale sample of five *W* responses [1, p. 217], and the subject's actual IQ was added to the scale card.

Procedure

In the study proper, each judge estimated the intelligence of 25 subjects. All that the judge saw for any one subject was five *W* responses selected from the subject's Rorschach according to an arbitrary system. Each sample of five *W* responses, including inquiry, was typed on a 5 × 8 card. The judge compared this card with the five scale sample cards, and placed the subject at one of the five points on the judgment scale.

Four experienced clinicians judged the 25 cases. For incidental comparison, four less experienced clinicians in training judged the same 25 cases.

The assigned scale values were used in calculating product-moment correlations for judgments based on Rorschach *W* responses with Wechsler IQ. The pooled judgments of two judges were correlated with those of the other two to obtain a reliability coefficient.

Results

The four experienced clinical psychologists made valid judgments of intelligence based on five Rorschach *W* responses for each subject. The correlations of their judgments against Wechsler IQ were .61, .62, .69, and .73. The mean correlation for any one judge, computed by Fisher's method of weighted z's [2, p. 363], was .68, with a standard error of .11. The reliability coefficient for two judges against two judges was .84.

The less experienced judges also made valid judgments of intelligence. For these less experienced judges, the four correlations were .45, .51, .54, and .59. It will be noted that the highest correlation for a less experienced judge was not as high as the lowest correlation for an experienced judge. However, the mean correlation for the less experienced judges, .54, was not significantly lower than the mean for experienced judges, .68. The reliability correlation for two less experienced judges against the other two was .69. All correlations were significant at the .01 level of confidence, except the lowest for a less experienced judge which was significant at the .04 level.

Because of the small *N* of 25, the validity rho of .77 for 10 cases obtained in constructing the judgment scale and reported in the section on *The Judgment Scale* is of interest. These 10 cases were additional cases not included in the 25 cases, and may be considered at least a suggestive replication.

Details of Method

The essential purpose of this small study was to illustrate and consider certain details of procedure in applying judgment scales to projective material. These details will be listed, with brief comment, in this section under the following headings: (*a*) rationale, (*b*) criterion, (*c*) material judged, (*d*) judges, (*e*) the judgment task, and (*f*) generalization from results.

Rationale. In general, studies such as this may be regarded as simplified checks on the projective approach and on specific claims for the significance of unstructured techniques. Eventually, a sufficient body of such studies should provide a sound basis for the clinical use of projective techniques. However, it seems necessary to emphasize repeatedly that the purpose is not to replace the whole projective approach with a set of scored scales which are a botched compromise, neither carefully thought

out in construction as traditional measuring scales, nor retaining the advantages of the projective approach.

A second consideration in terms of general purpose relates to the question of validation. In this study, we report a correlation between judgments of unstructured behavior and scores on a highly structured test. The two samples of behavior represent approaches from two markedly different origins. However, we are not validating projective material in the sense of predicting future behavior in a real-life situation. A number of studies of projective material have reached for predictive goals that are too far in advance of our present knowledge, and have suggested a misleading answer to such a simple question as: Will a projective technique correlate as well with an intelligence test as another intelligence test will? Also, in this study, we were chiefly interested in details of the judging process. Details of procedure can be considered best in a context where the correlation is at least a point or so above zero.

Criterion. The outside criterion selected for this study was the Wechsler-Bellevue IQ. The purpose of the study determined this selection, not the significance of the variable as one of the most important reflected in projective procedures.

We chose an intelligence test because, in the field of single artificial variable measurement, intelligence was the variable our judges knew most about. By training and experience, the judges had developed a definition of intelligence that was markedly influenced by the actual operations involved in the Wechsler. Every judge had used the Wechsler extensively with patients similar to the subjects of the study. They had had informal clinical experience, with correction of errors, in estimating Wechsler IQ from Rorschach performance. When they made their judgments in the study, they knew that the outside criterion was the Wechsler. The judges were as familiar with the judgment called for, and with the outside criterion, as they were with the projective material judged. Finally, the outside criterion had an adequate degree of known reliability and validity.

Material judged. For each subject in this study, the material judged consisted of five *W* responses given by the subject in a routine Rorschach. We wanted a sample of material supported by clinical experience as related to intelligence, and a sample retaining some degree of configurational complexity. However, we also wanted the sample to be small, representative for each subject, and comparable for each subject.

Clinical experience over a period of time has tended to support *W* quality as related to intelligence. It is also probable that five *W* responses will retain much of the configurational flavor of the total Rorschach. At the same time, the sample for each subject was brief and could be typed on a 5 × 8 card. The judges were never overwhelmed by a unit of material that was so complex as to be bewildering. The selection of *W* responses made possible the elimination of any subject who did not happen to give an adequate sample of behavior pertinent to the judgment required. The material judged was also comparable for each subject. The problems that arise when, for example, the same judgment is made on a Rorschach with 100 responses and on a Rorschach with 15 responses were avoided.

Judges. In planning the present study, the specific training and experience of available judges were given careful consideration. These judges knew the projective technique, were thoroughly acquainted with the outside criterion in actual operation, and had worked with patients similar to those in the study. Their specific experience was both extensive and recent. This attention to specific and recent experience may be more important than considerations of general academic background and clinical training.

The judgment task. The present study was planned to make the judgment task as simple as possible, yet meaningful in terms of the judges' clinical experience. This consideration accounts for the small number of cases, and for the small sample of responses per subject. The judges could work rapidly. They had no feeling that their judgment energy was exhausted after the first two cases. At the same time, the *W* responses seemed to be units that were sufficiently complex to retain an adequate degree of configurational meaning. The judges felt that they had at least some basis

for estimating intelligence.

The simple judgment scale was an aid rather than an encumbrance. Only five points were used on the scale. The five samples of behavior could be kept in mind. Each point was labeled with a known IQ as well as represented by a brief sample of *W* responses. Finally, misjudged samples had been eliminated from the scale. Since each clinician may make judgments on a different basis, falsely judged scale samples may be confusing.

Generalization. A final comment should be made in regard to generalizing from the results of such a study as this. Disregarding the small *N* and the desirability of replication, let us suppose that a given degree of correlation can be established between these judgments and Wechsler IQ. Are we then justified in saying that the Rorschach measures intelligence? or that other judgments made from the Rorschach are valid? These questions are obviously not answered. Possibly, however, with a sufficient body of studies, we may be able to set up general principles in regard to the conditions necessary and the limitations indicated for such judgments to be valid.

Summary

1. The study was planned with the specific purpose of illustrating and considering certain details of procedure in studies of judgment applied to unstructured material. Comment on some of these has been presented in the section *Details of Procedure.* The considerations discussed are related to the conditions necessary for valid judgments, and to the limitations indicated.

2. In the study, experienced clinical psychologists made reliable and valid judgments of intelligence based on five Rorschach *W* responses. The correlation with Wechsler-Bellevue IQ was .68; the reliability correlation, .84.

3. Less experienced psychologists in training also made reliable and valid judgments of intelligence on the same basis. Their judgments tended to be less reliable and valid, but not significantly so.

Received October 22, 1953.

References

1. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
2. Kelley, T. L. *Fundamentals of statistics.* Cambridge, Mass.: Harvard Univer. Press, 1947.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

The Clinician as Judge: Judgments of Adjustment from Rorschach Single-Card Performance[1]

S. Thomas Cummings
Veterans Administration[2] and University of Pittsburgh

A review of the literature in clinical psychology in the past two decades reveals an increasing interest in the process and results of clinical judgment. Among others, Murray [10] and Hunt [8] have long urged a more intensive focus on clinical judgment and its ingredients—Murray in calling attention to "psychology's forgotten instrument," the interpreter of test data, and Hunt in his consistent emphasis on the need to validate the clinician.

A strong impetus to investigation of clinical judgment stems from the rapid expansion of the field of projective techniques with their inherent validation problems. The data yielded by projective techniques, while for the most part qualitatively wealthy and clinically useful, have offered many vexing problems in categorization and quantification leading toward validation. Validation efforts based primarily on the use of isolated scoring abstractions, as for example the number of human movement responses in a Rorschach, do not represent an adequate replication of the essential operations involved in making inferences about personality from such techniques, nor have the validating efforts based on such abstractions been highly productive of significant relationships with personality variables. On the other hand, research with projective technique data which more closely approximates the operations of the clinician-in-action, i.e., emphasizes the use of clinical judgment, has often, although not uniformly, yielded significant relationships with such variables [1, 2, 3, 4, 5, 11, 12, 13]. Such variables as the complexity of the judgment unit and the definability and stability of criteria will limit research in this area, as Hamlin [7] has emphasized in his review of variables in clinical judgment research.

The experiment reported here evaluates the reliability and validity of judgments of adjustment inferred from single-card performance on the Rorschach test. It was designed to permit maximum use of clinical judgment in evaluating the test data, while presenting judges with a circumscribed judgment task, the performance of a subject on a single Rorschach card.

Method

This research was conducted in two phases: a scale construction phase and a scale evaluation phase. In the scale construction phase a ten-point adjustment scale was constructed for each of the 10 Rorschach cards. In the scale evaluation phase, the ability of judges using these scales to discriminate between the Rorschach single-card performances of subjects of varying adjustment status was evaluated. The reliability and validity figures reported here refer exclusively to the judgments of adjustment made in the scale evaluation phase.

Scale construction phase. In the first phase, a ten-point Rorschach scale of adjustment was constructed for each Rorschach card, based on the pooled preferences of six judges who judged the single-card Rorschach performances of the 10 scale construction subjects in terms of adjustment. The scale construction subjects were white male veterans of World War II who showed limited variation in age and ed-

[1]Read at the meetings of the Eastern Psychological Association, Boston, April, 1953. Part of a dissertation submitted to the University of Pittsburgh, 1950. The author wishes to express his appreciation for the guidance received from his dissertation committee members, Drs. George W. Albee, Roy M. Hamlin, Gerald R. Pascal, and Carroll A. Whitmer.

[2]From the Veterans Administration Hospital, Aspinwall, Pa.

ucational characteristics, but extensive variation in adjustment status. In terms of nosology, the following gross categories were represented in the adjustment continuum: one normal (a nonvolunteer college undergraduate), five outpatient neurotics, two outpatient schizophrenics, and two hospitalized schizophrenics. The patients represented were chosen by the experimenter on the basis of case history information without reference to Rorschach data.

The six scale construction judges judged the single-card Rorschach performances of the 10 scale construction subjects using the method of paired comparisons. Typically, a judge was presented with 45 folders containing the pairing of each subject with every other subject on a Rorschach card and asked to indicate for each pair, "Which Rorschach performance indicates the better adjusted individual?" Wherever "adjustment" required definition in this research, as in instructions to judges, the following combination of characteristics was referred to them as defining adjustment: "ability to get along with other people, psychosexual maturity, freedom from anxiety and psychopathology, and perception of reality in accord with prevailing cultural norms."

The quantitative values assigned to the ten points on each of the adjustment scales were derived by the mean-proportion method for handling paired-comparison data, described by Guilford [6]. This method converts the proportion of total preferences awarded to each subject, out of the total possible number of preferences on a Rorschach card, into normal curve base line deviates. By this technique, considerable variability in the range and spacing of scale values is possible, based on the relative consistency of the judges in their preferences.

These procedures yielded a ten-point scale of adjustment for each Rorschach card, consisting of the ranked performances of the 10 subjects on each card and the scale values derived for them.

Scale evaluation phase. In this phase of the research, evaluation was made of the ability of judges using these adjustment scales to discriminate between the Rorschach single-card performances of members of five diagnostic groups, representing different levels of adjustment.

A sample of 50 scale evaluation subjects was assembled with 10 members in each of the five following groups: Group I, normals; Group II, nonhospitalized neurotics; Group III, hospitalized neurotics; Group IV, nonhospitalized schizophrenics; Group V, hospitalized schizophrenics. The normals in the sample were a group of nonvolunteer college undergraduates. The patients were gathered from two outpatient clinics and two hospitals by selecting consecutive cases that satisfied the sample's qualifying characteristics from the current files of these services. The general qualifying characteristics for inclusion in this sample were as follows: male, white, 20-55 age range, one-half year or more of college training, no suspicion of having intracranial pathology. Additional qualifying criteria were imposed for admission to subgroups, regarding length of hospitalization or nonhospitalization and manner in which diagnosis was reached [4].

It was recognized from the outset that definition of the independent variable, adjustment status, in terms of membership in these five diagnostic groups was a gross definition that would tend to attenuate the relationships evaluated [9]. However, such a gross criterion definition is frequently used in prevailing practice, and it was our opinion that if results suggestive of significance were obtained with such a criterion, further refinement of the criterion would lead to enhancement of the significance of these relationships.

The scale evaluation judgments of adjustment were completed by two judges, neither of whom had participated in the scale construction phase. These judges made their judgments of adjustment in the following fashion. The 10 single-card performances representing an adjustment scale for the Rorschach card in question were ranked before the judge with the instruction that these performances represented an adjustment continuum achieved by prior rating. He was then asked to familiarize himself with the adjustment levels represented by the scale members and, following this, to make his judgments of adjustment of the 50 scale evaluation subjects' single-card performances by allocating each to the appropriate point on the scale, in terms of its being most closely com-

parable with one of the scale members in adjustment level.

The ranks assigned to the 50 scale evaluation subjects by these judges using the scales were converted into predetermined scale values. Reliability coefficients were computed for the judgments of adjustment made on each card and for total series of judgments by product-moment correlation of these scale values. Spearman-Brown correction of these coefficients produced measures of these judges' joint reliability.

The validity of these judgments was evaluated primarily by determining the degrees of association between Rorschach adjustment scale values and adjustment group status, i.e., membership in one of the five diagnostic groups. In evaluating the degrees of association between these variables, the unbiased correlation ratio, epsilon, corrected for broad categories, was computed for individual cards and for the total series, i.e., the average score for all cards combined. A further refinement of the validity evaluation procedures which was made was evaluation of the significance of differences in total series mean adjustment score, using the *t* test. In all these computations, the average of the two scale values assigned by the two judges to a subject's Rorschach performance constituted his adjustment score for that card. "Total series mean adjustment score" either for individual subjects or for adjustment groups is the average of the scores achieved on all ten Rorschach cards.

Results and Discussion

Table 1 presents the reliability and validity coefficients for each Rorschach card and for the total series achieved by the scale evaluation judges.

Reliability. The judgments of adjustment made by the scale evaluation judges showed high degrees of reliability. Their reliability coefficient for the total Rorschach series was .89, a coefficient significant at the 1 per cent level. On individual cards their reliability coefficients ranged from .90 on Card III to .74 on Card V, all significant at the 1 per cent level of confidence. Despite the appreciable spread in values between the extremes of this series, statistical evaluation of the significance of intercard differences in reliability suggests that the judges showed little true differential reliability in making their judgments from card to card. Only comparison of the two highest coefficients on Cards III and IV with the lowest on Card V and Card III with the next lowest on Card I yielded ratios suggestive of significance at the 5 per cent level. With reference to consistency of judgment then, these judges showed high degrees of reliability in making their judgments of adjustment.

Table 1

Reliability and Validity Indices of Scale Evaluation Judges' Judgments of Adjustment

Card	Reliability		Validity	
	r_c	p	ε	p
I	.78	.01	.29	
II	.86	.01	.46	.01
III	.90	.01	.50	.01
IV	.89	.01	.36	.05
V	.74	.01	.44	.01
VI	.86	.01	.41	.05
VII	.80	.01	.61	.01
VIII	.85	.01	.25	
IX	.82	.01	.40	.05
X	.86	.01	.43	.05
Total series	.89	.01	.50	.01

Validity. The validity of these judgments was first evaluated by correlating mean adjustment score and adjustment group status. The epsilon value between mean adjustment score from the total series and adjustment group status was .50, significant at the 1 per cent level. The epsilon values on individual cards ranged from .61 on Card VII to .25 on Card VIII, as reference to Table 1 will illustrate. The four highest coefficients on individual cards are significant at the 1 per cent level, the next four at the 5 per cent level, and the lowest two fall below the 5 per cent level. Thus, these judges' judgments of adjustment showed medium-order and statistically significant degrees of association with the criterion, diagnostic status.

Further evaluation of the ability of these judges to discriminate between the adjustment levels of these subjects was made by comparing differences in mean adjustment score between the five adjustment groups. Table 2 summarizes the sample's statistics for total series mean adjustment score. Inspection of these group means shows a consistent decline from Group I to Group V, albeit some of the differences in adjacent means are small. Evalua-

Table 2
Comparison of Adjustment Groups in Total Series Mean Adjustment Score

Statistic	Adjustment group				
	I	II	III	IV	V
N	10	10	10	10	10
Mean	1.33	1.30	1.13	1.05	.81
SD	.10	.15	.28	.15	.43
SE_m	.03	.05	.09	.05	.14

tion of the significance of differences in these group means is presented in Table 3. Here, 5 of a possible 10 significant differences result. The normals (Group I) are significantly higher than the bottom three groups and the nonhospitalized neurotics (Group II) significantly higher than the bottom two groups. Three other fairly sizable differences between means fall short of statistical significance because they involve relatively high variances within this setting of low degrees of freedom.

Table 3
Fisher *t* Ratios on Adjustment Group Differences in Total Series Mean Adjustment Score

Group	II	III	IV	V
I	.50	2.22*	4.67**	3.71**
II		1.70	3.57**	3.27**
III			.80	1.88
IV				1.60

* Significant at 5% level of confidence.
** Significant at 1% level of confidence.

In essence, then, the judges' judgments of adjustment of these 50 subjects' single-card Rorschach performances show strong positive relationships with the external criterion of adjustment used, diagnostic status. Admittedly, the size of the relationships found has little value for purposes of individual prediction but leaves little doubt as to the mutual implication of these variables, particularly in view of the vagueness of the adjustment concept and the grossness of the criterion.

The essentially positive results obtained in this research invite comparison with two similar studies, one by Grant, Ives, and Ranzoni [5] and the other by Newton [11], both of which related judgments of adjustment from the Rorschach to external criteria of adjustment and yielded negligible relationships. Both studies used total Rorschach protocols as the judgment unit. The former used a simple descriptive four-point scale of adjustment, and case-worker ratings of adjustment based on interviews as a validating criterion. The latter used a seven-point adjustment scale derived from the research reported here; ranked total Rorschachs illustrated the points on the scale; the external criterion was psychiatric judgments of case history information, which, incidentally, correlated highly (.79) with diagnostic status, the criterion used in the research reported here. Without evaluating other attenuating factors, which these authors review comprehensively, we would emphasize with Hamlin [7] that complexity of judgment unit, i.e., total Rorschachs, was a common factor in these studies which probably tended to diminish the obtained relationships. With 10 judgments of a less complex judgment unit, the single card, it would appear that a judge's opportunities for intensive coverage of his data are greatly enhanced. Even though the factor of intercard variability in performance is sacrificed by this method, the greater familiarity with one's data afforded by it through making 10 judgments of coherent parts of a whole would seem to more than compensate for this.

A related methodological factor in clinical judgment studies that has received little attention but probably also has an attenuating effect is similar to the factor of "esthetic fatigue" reported in esthetic judgment studies, something we could call "rater demoralization and fatigue." This appears to be primarily a function of complexity of judgment unit and size of total judgment task. Many judges who participate in clinical judgment studies report feelings of dissatisfaction and of being overwhelmed by having to retain and manipulate so many complex variables simultaneously and having to continue to do so for what is perceived as a mountainous mass of protocols. If such a factor exists, it is probable that it would have a further depressing effect on relationships obtained.

Finally, worthy of restatement, despite its obviousness, is the point that this and similar studies have limited implications for the use of the Rorschach or any test as a clinical tool. Inferences gained from a test in clinical use are typically evaluated within a context of other test and interview behavior data, frequently including health and historical information. Blind interpretation of tests is a worth-while

training device, an interesting pastime, and a useful methodological feature of research on the bases and limits of clinical judgment. However, it provides crucial validation information chiefly in evaluating the assertions of single-test partisans regarding the use of their favorite techniques in isolation, not in evaluating the validity of these techniques as they are characteristically used in clinical practice.

Summary

A ten-point Rorschach scale of adjustment was constructed for each Rorschach card. Six judges used the method of paired comparisons in judging the single-card Rorschach performances of 10 subjects representing a broad range of adjustment. The 10 ranked single-card Rorschach performances plus the scale values derived for them comprised a scale of adjustment.

Two additional judges used these scales in evaluating the adjustment levels implicit in the Rorschach performances of 50 subjects of varying adjustment status (five groups of 10 subjects each of normals, nonhospitalized neurotics, hospitalized neurotics, nonhospitalized schizophrenics, and hospitalized schizophrenics).

These judges made judgments of adjustment that were highly reliable. The validity of their judgments as related to the criterion, diagnostic status, was for the most part of medium order, not permitting accurate prediction of diagnostic status from knowledge of mean adjustment score (average of judges' ratings) alone, but indicating statistically significant relationships between these variables. The indeterminate unreliability in the criterion was noted as probably attenuating the relationships obtained here.

Certain factors in clinical judgment research, such as complexity of judgment unit and "rater demoralization and fatigue," are briefly discussed.

Received November 20, 1953.

References

1. Albee, G. W., & Hamlin, R. M. An investigation of the reliability and validity of judgments of adjustment inferred from drawings. *J. clin. Psychol.*, 1949, **5**, 396-398.
2. Benjamin, J. D., & Ebaugh, F. G. The diagnostic validity of the Rorschach test. *Amer. J. Psychiat.*, 1938, **94**, 1163-1178.
3. Bialick, I. A method of applying clinical judgment to problems of Rorschach validation: the quality of Rorschach whole responses and intelligence. Unpublished master's thesis, Univer. of Pittsburgh, 1950.
4. Cummings, S. T. An investigation of the reliability and validity of judgments of adjustment inferred from Rorschach test performance. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1950.
5. Grant, Marguerite Q., Ives, Virginia, & Ranzoni, Jane H. Reliability and validity of judges' ratings of adjustment on the Rorschach. *Psychol. Monogr.*, 1952, **66**, No. 2 (Whole No. 334).
6. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
7. Hamlin, R. M. The clinician as judge: implications of a series of studies. *J. consult. Psychol.*, 1954, **18**, 233-238.
8. Hunt, W. A. The future of diagnostic testing in clinical psychology. *J. clin. Psychol.*, 1946, **2**, 311-317.
9. Masserman, J. H. Diagnosis and prognosis in psychiatry. *J. ment. Sci.*, 1938, **84**, 893-946.
10. Murray, H. A. *Manual for the Thematic Apperception Test.* Cambridge, Mass.: Harvard Univer. Press, 1943.
11. Newton, R. L. The clinician as judge: total Rorschachs and clinical case material. *J. consult Psychol.*, 1954, **18**, 248-250.
12. Vernon, P. E. The significance of the Rorschach test. *Brit. J. Med. Psychol.*, 1935, **15**, 199-217.
13. Vernon, P. E. Recent work on the Rorschach test. *J. ment. Sci.*, 1935, **81**, 894-920.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

The Clinician as Judge: Total Rorschachs and Clinical Case Material[1]

Richard L. Newton
University of Pittsburgh

This research is an investigation of the value of a clinical technique (the Rorschach test) and the validity of clinical judgment as reflected in interpretations of total Rorschachs and clinical case material.

The study had three objectives. The first and primary objective was to ascertain to what degree psychologists' judgments of adjustment from Rorschach protocols agree with psychiatrists' judgments of adjustment from clinical material on the same subjects.

The second objective was to discover to what extent Rorschach judgments agree with psychologists' judgments of clinical case material. The third objective was to determine to what degree psychiatrists and psychologists agree in making judgments of clinical case material.

Method

Adjustment Scales

As a preliminary step, two quantitative scales of adjustment were constructed. Cummings [2] derived a psychiatric adjustment scale by having each of six qualified psychiatrists rank a group of 10 subjects in terms of over-all adjustment. The six psychiatrists were presented with abstracts compiled from the case histories of the subjects. The 10 subjects represented a range of adjustment from social adequacy to psychosis necessitating hospitalization. Normalized scale values were computed from mean ranks to construct a ten-point scale of adjustment. Cummings estimated the reliability of the psychiatric scale rankings for these 10 subjects by a split-judging group technique, which yielded the highly significant product-moment correlation of .95. In the present study, Cummings' scale was reduced to seven points to eliminate tied values and to produce more discrete spacing of the scale values. The elimination of three scale values did not narrow the range of adjustment represented. In the remaining seven subjects there were one socially adequate subject, three neurotics, one nonhospitalized schizophrenic, and two hospitalized schizophrenics. All were white male veterans of World War II.

A Rorschach adjustment scale was derived by having each of six qualified Rorschach workers rank the complete Rorschach protocols of these seven subjects in order of over-all adjustment. Normalized scale values were computed from mean ranks, producing a seven-point Rorschach adjustment scale. The reliability of the Rorschach rankings for the seven subjects was estimated by a split-judging group technique, which yielded the highly significant product-moment correlation of .94.

The Use of the Adjustment Scales

The scales for clinical histories and total Rorschachs were employed in obtaining judgments of the over-all adjustment of each of 50 experimental subjects made up of five groups of 10 subjects each. The five groups were socially adequate subjects, nonhospitalized neurotics, nonhospitalized schizophrenics, hospitalized neurotics, and hospitalized schizophrenics. The sample of 50 subjects was comparable in age, education, sex, color, and adjustment range to the seven subjects employed in constructing the adjustment scales.

Psychiatric adjustment scale. Each of 10 qualified psychiatrists was asked to judge abstracts of clinical material (basic data sheets) of 15 of the experimental subjects in terms of adjustment as represented by the range covered by the seven scale subjects. Each judge was

[1]This article is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree, University of Pittsburgh, 1952.

presented with the basic data sheets of the seven scale subjects and used the relative scale placements of these subjects as a guide in placing each of the 15 experimental subjects at a point on the seven-point adjustment scale. Thus, three independent judgments of adjustment were obtained for each experimental subject; these made possible the computation of an estimate of the reliability of psychiatric judgments.

A group of 10 clinical psychologists also made judgments of adjustment from the basic data sheets of the experimental subjects in exactly the same manner as that employed by the 10 psychiatrists.

Rorschach adjustment scale. The same 10 psychologists, all skilled in the Rorschach, made judgments of adjustment for the experimental subjects from global Rorschachs. Each Rorschach judge was presented with the Rorschachs of the seven scale subjects and used the relative scale placements of these Rorschachs as a guide in placing each of 15 of the experimental Rorschachs at a point on the seven-point Rorschach adjustment scale. None of the Rorschachs judged by any Rorschach expert was obtained from subjects whose basic data sheets he had judged. Therefore, there was no possibility of a psychologist associating a subject's basic data sheet with that subject's Rorschach.

Results

The following results were obtained:

1. Psychiatrists' judgments of basic data sheets revealed a very high degree of reliability (.91).

2. Psychologists' judgments of Rorschachs revealed a high degree of reliability (.73).

3. Psychologists' judgments of basic data sheets revealed a very high degree of reliability (.94).

4. Psychologists' judgments of Rorschachs were compared with psychiatrists' judgments of basic data sheets to determine the degree of relationship or agreement between the judgments of the two disciplines when each discipline used a different instrument of evaluation. The resulting correlation coefficient (.09) was not significantly different from zero.

5. Psychologists' judgments of Rorschachs were compared with psychologists' judgments of basic data sheets to determine the degree of relationship between the two sets of judgments made by the same discipline with two different instruments of evaluation. The resulting correlation coefficient (.20) was not significantly different from zero.

6. Psychologists' judgments of basic data sheets were compared with psychiatrists' judgments of basic data sheets to determine the degree of relationship between the judgments of disciplines when both disciplines were using the same instrument of evaluation. The resulting correlation coefficient (.86) is significant.

7. Finally, psychologists' judgments of Rorschachs were compared with placement in the five nosological groups to determine to what degree the Rorschach judgments were associated with diagnostic categories. The resulting eta (.23) was not significantly different from zero.

Discussion

The question arises as to whether or not the criterion was a good one. The following evidence suggests that the criterion was good. Very high reliability resulted from the judgments of the criterion material by judges of both disciplines, thus showing agreement within each of the disciplines. The correlation (.86) between disciplines in judging the criterion material was very high also, showing general agreement among the "experts."

Newton and Goodman [4] compared the psychiatrists' judgments of the basic data sheets with placement in five diagnostic categories and obtained an eta of .79, significant at the 1 per cent level. These findings suggest that the criterion of psychiatric judgment is positively correlated with diagnostic classification. However, in the same study, Newton and Goodman found that psychiatrists' judgments of basic data sheets did not correlate significantly with Buhler Basic Rorschach Scores [1] of the 50 Rorschachs. Although their study constituted a different approach to Rorschach evaluation, the findings were consistent with those of the present study. The findings of Hamlin and Newton [3] suggest that concentrated study of limited data will produce positive results.

Summary

This research investigated the validity of

clinical judgments of adjustment based on global interpretations of Rorschachs and on clinical case material. The method employed incorporated an adaptation of psychophysical scaling techniques wherein a Rorschach scale of adjustment was derived from psychologists' judgments of adjustment of Rorschach protocols and a psychiatric adjustment scale was derived from psychiatrists' judgments of adjustment based on clinical case material.

The use of these scales by psychologists and psychiatrists in judging the adjustment of 50 experimental subjects resulted in four basic findings:

1. High reliability was obtained in each of the three judging tasks: clinical case material by psychologists, by psychiatrists; Rorschachs by psychologists.

2. A significant relationship was obtained between psychologists' and psychiatrists' judgments of case material.

3. Rorschach judgments did not correlate significantly with psychologists' judgments of case material or with psychiatrists' judgments of case material.

4. Rorschach judgments did not correlate significantly with placements in five diagnostic categories.

Received November 18, 1953.

References

1. Buhler, Charlotte, Buhler, K., & Lefever, D. W. *Development of the basic Rorschach score with manual of directions.* Los Angeles, Calif.: Rorschach Standardization Studies, No. 1, 1948.
2. Cummings, S. T. The clinician as judge: judgments of adjustment from Rorschach single-card performance. *J. consult. Psychol.*, 1954, **18**, 243-247.
3. Hamlin, R. M., & Newton, R. L. Comparison of a schizophrenic and a normal subject, both rated by clinicians as well-adjusted on the basis of "blind" Rorschachs. Paper read at Eastern Psychol. Ass., Atlantic City, N. J., March, 1952.
4. Newton, R. L., & Goodman, H. W. A comparison of Basic Rorschach Scores with judgments of adjustment based on clinical case material. Paper read at Eastern Psychol. Ass., Atlantic City, N. J., March, 1952.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Intra-Individual Consistency in Response to Tests of Varying Degrees of Ambiguity[1]

Arthur C. Carr
Creedmoor State Hospital and Adelphi College

An assumption implicit in all testing procedures is that the behavior elicited in the test is a representative sample of the behavior characteristic of the individual in other situations as well [29]. Projective techniques in particular lean heavily upon the assumption that all behavior manifestations, the seemingly superficial as well as the obvious, are expressive of the individual's personality [1, 17, 24].

Recent research, however, would seem to justify the caution of Bellak that "the projective hypothesis . . . has to be handled with considerable care" [1, p.5]. The literature is replete with findings which suggest that behavioral measures based on projective techniques may be inconsistent and unpredictable [3, 5, 7, 11, 16, 22, 25, 26, 27].

In clinical practice an assumption nevertheless is held that apparently unrelated intra-individual responses may be motivationally identical, i. e., may stem from the same causal structure. In light of this assumption, it is well to remember that tests merely record *behavior*, while clinical insights are arrived at through *inference* [17]. It follows that consistency or orderliness in personality must be found in what is inferred to lie behind the responses of the individual rather than in the varied overt aspects of his behavior.

The purpose of the present study was to investigate an aspect of the projective hypothesis dealing with the consistency of an individual's responses to projective tests that vary as to degree of ambiguity or structuredness. It seemed obvious from previous considerations that any meaningful consistency in behavioral expression must be found in what is "behind," or in what motivates, a given response rather than in some gross or surface feature of the response. Therefore, the diverse test responses were organized in terms of what could be inferred to be their underlying feeling tone, emotion, attitude, or motivating force. This quality was called the "inferred affect" of a response. In terms of relevance for personality dynamics, a classification of four fundamental qualities evolved: *hostility*, *positive feelings*, *anxiety*, and *dependency*.

The investigation studied the intra-individual consistency of responses elicited by the Sentence Completion Test (SCT), the Thematic Apperception Test (TAT), and the Rorschach test. The Forer [12] version of the SCT was chosen because of its systematic coverage of interpersonal relationships.

As is apparent in Table 1, each of the chosen tests contains stimuli that are assumed to elicit responses from which can be inferred the subject's attitude or feeling tone with respect to important human figures—mother, father, females, males, people, and persons in authority. These stimuli vary in ambiguity—from

[1]Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are a result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

Adapted from a dissertation submitted in partial fullillment of the requirements for the Ph. D. degree, University of Chicago, 1952. The author would like to thank Dr. Donald W. Fiske for his kind assistance and understanding guidance throughout all phases of the research on which this article is based, as well as Drs. Hedda Bolgar and Julius Seeman, other committee members. The author is also grateful to those who served in varying capacities in scoring test data, particularly Mr. Austin Herschberger, who was especially generous with both time and clinical skills.

Table 1
Test Stimuli Pertaining to Various Interpersonal Figures

Figure	Forer SCT	TAT	Rorschach
Mother	20. His earliest memory of his mother was 35. Most mothers 60. My mother 76. When my mother came home, I 94. When he was with his mother, he felt 99. Mothers	Picture 6 BM (Elderly woman and young man)	Card VII "Mother" card
Father	9. His father always 16. Most fathers 33. My earliest memory of my father 70. When my father came home, I 88. Whenever he was with his father, he felt 96. Fathers	Picture 7 BM (Elderly man and younger man)	Card IV "Father" card
Females	21. When I meet a woman, I 26. Sisters 39. Most women 67. My first reaction to her was 71. As she spoke to him, he 81. Most women act as though	Picture 4 (Woman clutching man's shoulders)	Female percepts seen (Cards I-X)
Males	10. Men 13. My first reaction to him was 30. Most men 50. Brothers 59. While he was speaking to me, I 95. Most men act as though	Picture 9 BM (Men lying on grass)	Male percepts seen (Cards I-X)
People	15. His new neighbors were 38. I feel that people 49. A lot of people 66. When I meet people I generally feel 84. Most people are 93. Whenever he is introduced to people, he	Picture 2 (Farm scene)	Human percepts Animal percepts (Cards I-X)
Authority	23. When he met his boss he 36. Taking orders 73. People in authority are 77. When they told him what to do, he 91. Sometimes I feel that my boss	Picture 1 (Boy and violin)	Card IV

statements on the SCT directly referring to such people as "my mother," to stimuli of comparable but less obvious meaning, as the cloudlike Rorschach inkblot which has also been assumed by psychologists to represent the mother figure [21, 23]. On such an implied continuum, the TAT picture of the older woman and the young man (the "mother-son" card) would lie between the extremes. Such variations in stimuli are also present for other important figures. For example, the SCT includes items referring directly to such figures as mother, father, females, males, people, and authority figures, while the Rorschach and TAT provide means for assaying the same relationships through stimuli whose representa-

tional values are less immediately apparent.

The study was concerned primarily with the intra-individual consistency of responses elicited by varying stimulus representation of *each* interpersonal figure. It is to be stressed that these responses were classified in terms of their inferred affect, i.e., hostility, positive feelings, anxiety, and dependency.

A frequent assumption in clinical practice is that stimuli more similar to the object being considered will elicit tendencies closer to awareness than will stimuli that are less like the object. It is assumed that affect and impulses denied expression toward the real object, or any close representation of it, are nevertheless elicited by less "threatening" stimuli which, by virtue of certain qualities, still represent or "stand for" that object. Although Lazarus [18] has recently suggested the desirability of an even further extension of the ambiguity continuum than that represented in the usual battery of diagnostic tests, the tests used in this study represent a wide range on such a distance continuum. Therefore, it was believed that the resulting analysis would provide a comparison between the covert and the more overt aspects of personality functioning and reveal any disparity in such levels of function.

Method

The projective tests were administered to fifty male patients in a mental hygiene clinic. Formal diagnostic classifications undoubtedly would have revealed a range from mild neurosis to that of borderline psychosis. In spite of the differences in levels of adjustment, however, all subjects were functioning with some degree of success in everyday interpersonal relationships. No attempt was made to control the group for intellectual level, age, or diagnostic classification. It was believed that these factors within any one individual would be equally rather than differentially operative in all tests. The subjects ranged in age from 24 years to 41 years, with a mean of 32. Intelligence quotients (W-B) ranged from 84 to 147, with a mean of 111.

The tests were administered in the normal course of routine psychological testing. The entire 100 items of the SCT were administered to each subject. The pertinent items were later abstracted for purposes of this research. The Rorchach test was administered according to the methods suggested by Beck [2]. Seven cards were chosen from the TAT on the basis of the applicability of each card to assay the interpersonal relationship under consideration (Table 1). Three qualified psychologists chose these cards as most appropriate for eliciting affect toward each specified figure. Administration followed standard TAT procedure.

The inferred affect of all pertinent responses was independently scored by two qualified clinicians. Each test was independently scored, neither judge being aware of which response belonged to any subject.

It became obvious that the original intention of maintaining similar categories of inferred affect within and between tests could not be sustained, because all stimuli did not evoke productions whose qualities were characteristic of every anticipated category. Thus, while the SCT evoked a large proportion of positively toned responses, the only qualities inferred from Rorschach responses proved to be those of hostility, anxiety, and dependency, with a category of neutral or unclear responses. Similarly, the categories within a given test were not applicable to all the various figures. For example, no dependency category was inferred from TAT themes pertaining to females, males, people, or authority figures. The categories of inferred affect, therefore, were ultimately established in terms of the particular set of responses evoked from the particular group of subjects.

Rorschach scoring. After the Rorschach tests had been scored according to standard procedures established by Beck [2], every human and animal percept (*H, Hd, A,* and *Ad*), along with its accompanying inquiry, was extracted from the total record and entered on coded cards. The first responses to Cards IV (father card) and VII (mother card) were also entered on separate cards. This procedure was used, rather than attempting to arrive at scores based on over-all reaction to the card, as research had suggested the reliability of this method [4]. Judges rated each extracted response for inferred affect in terms of one of the following qualities: hostility, positive feelings,

dependency, anxiety, and "other" (neutral or unclear). Through joint discussion of the criteria established by DeVos [9] and Elizur [10], they were acquainted with the general method for scoring. Although the numerous subcategories established by DeVos were not utilized, his quantitative approach to the scoring of Rorschach responses in terms of what he calls "affective symbolism" greatly facilitated the ordering of the data.

Of 662 responses so classified, judges agreed on 617, a percentage of agreement of 93.2. Percentage of agreement between judges was compared with that which would have been obtained on a chance basis. Critical ratios obtained on Rorschach scoring, as well as on other tests, indicated that on all comparisons the agreement obtained was significant at the .001 level. Differences in scoring were later resolved by discussion between the two judges. Where varying interpretations were not resolved (approximately 2 per cent of the total number), responses were necessarily placed in the "other" category.

TAT scoring. Prior to scoring the TAT, each story was extracted from the total record and coded so as to make later identification possible. It became obvious that a distinction had to be made between the theme's inferred affect and that of the characteristics overtly attributed to the test representative of the given interpersonal figure, since the qualities therein were sometimes dissimilar. For example, the theme of a TAT story involving the death of a kind, loving person would be scored *hostility* in contrast to the score based on the characteristics overtly attributed to the figure.

Overtly attributed characteristics (TAT-overt) were classified into two categories: positive feeling and hostility. This categorization was applicable to all productions except those elicited by Card 1, wherein an additional category was devised for those stories in which authority figures were not introduced.

The following criteria provided a basis for scoring TAT themes. These criteria pertain only to the quality inferred from the theme and are thus independent of the described characteristics of the person or persons in any TAT production:

Hostility

1. Pressure from partner or peer — prohibiting, compelling, punishing, quarreling, interfering.
2. Physical harm inflicted by peer or partner.
3. Physical harm inflicted to peer or partner.
4. Death or illness of peer or partner.
5. Disappointment in peer or partner.
6. Bad news to peer or partner.
7. Degradation or depreciation of peer or partner.
8. Departure — when they suggest conflicting goals, desire of hero to get away.

Dependency

1. Succor from peer or partner — seeking or receiving aid, advice, consolation, etc.
2. Nurture — bestowing aid, support, care.
3. Filial obligation — when duty and dependency predominate. If obligation implies hostility, this latter quality may predominate.
4. Departure — when departure is seen by hero as necessary to "becoming a man," etc.
5. Rebellion against authority, domination, dependency — when strivings for independence seem greater than hostile elements.

Anxiety

1. Concern over the future, outcome of situation.
2. Fear of what will happen, fear of the unknown.
3. Decision — where a difficult choice is posed.
4. Occupational concern.
5. Curiosity.
6. Moral struggle, illicit sex, guilt.
7. Departure — when departure is source of anxiety and grief to the hero.
8. Confession, admission of guilt.

In a total of 300 stories so rated, agreement between judges was 91 per cent on rating inferred effect of themes, and 94 per cent on overtly attributed characteristics of the figure being assayed.

SCT scoring. The original four categories of inferred affect were maintained in SCT scoring results—positive feelings, hostility, dependency, and anxiety. Neutral or unclear responses were categorized as "other." In general, scoring was done in terms of the obvious meaning of the completions, that is , taking completions at their "face value." Thus, "Most mothers *are good*" was scored *positive feelings* without questioning whether or not the real meaning was that "most mothers are good but mine is not," etc.

The category *positive feelings* pertained to all completions in which the primary attitude expressed toward the figure was that of acceptance, love, or security.

The category *dependency* included those attitudes which were primarily ones of dependency or immature compliance. Childishly toned responses, often pertaining to eating or being fed, or the expression of child-like feelings were classified in this category.

The category of *anxiety* was applied to all completions in which the primary attitude was that of fear, guilt, anxiety, or insecurity.

The category of *hostility* contained all those completions in which the primary attitude expressed toward the figure was that of avoidance, derogation, rejection, or pure hostility.

All other remaining responses which could not be fitted into the above categories were classed as *other*. Factual statements generally fell into this category.

A total of 35 sentence completions were scored for each subject. Agreement between judges was 1,594 out of 1,750 responses, or 91.1 per cent. Differences in scoring were resolved in discussion.

Statistical Treatment of Data

Ratings for each figure being assayed in each test were now available for every subject. Where the number of opportunities for such ratings varied (as with the number of female, male, human, and animal responses evoked by Rorschach stimuli), responses were equated by changing them into percentage of total responses [8]. Where the number was constant for all subjects (as with the completions for each figure on the SCT), occurrences of a given affect for each figure were simply totaled. Ratings for the total group of subjects were dichotomized in terms of the median point for number of occurrences (or presence and absence) of each affect category.

Each subject was now scored in terms of "high" and "low" placement in each affect category for each interpersonal figure. The tabulation compared "high" and "low" groups on each variable with the "high" and "low" group of every other variable.

Chi-square tests of independence were used to test the relationships between these variables [19]. Yates's correction for continuity was applied in all instances. One-tailed tests of significance were applied to all predicted relationships, i.e., the tabled probabilities for the chi-square distribution were halved because only positive relationships were to be considered as significant [15].

In the results presented, only these predicted relationships are dealt with. Ancillary findings are to form the basis of a separate report. The number of relationships predicted includes only those in which similar affect categories were maintained. Lack of a dependency category on certain figures assayed in the TAT, for example, eliminated the possibility of a consistency between it and dependency as elicited by the other tests' representations of the same figure. Since no similar category was present on which to have made a prediction of consistency, none was made. One could argue, however, that if real consistency prevailed, affect categories would necessarily have to be similar between all tests. Any dissimilarity between such categories should then be considered an exception to any basic consistency in personality expression. On the other hand, if a particular stimulus fails completely to elicit a certain affect category from a wide range of subjects, the evidence hardly should be interpreted as showing a lack of consistency within the subjects. It seems better to assume that the stimulus was structured in such a way (in spite of its relative unstructuredness) that cues

Table 2

Consistency of Inferred Affect for Predicted Relationships between Tests

Tests compared	Total number of relationships		
	Predicted	Significant at 5% level	Significant at 10% level
Rorschach and total TAT	22	6 ($p < .001$)	12 ($p < .001$)
Rorschach and TAT-theme	14	5 ($p < .001$)	9 ($p < .001$)
Rorschach and TAT-overt	8	1 (nonsig.)	3 ($p < .05$)
Rorschach and SCT	22	1 (nonsig.)	5 ($p < .05$)
Total TAT and SCT	27	9 ($p < .001$)	12 ($p < .001$)
TAT-theme and SCT	13	3 ($p < .05$)	4 ($p < .05$)
TAT-overt and SCT	14	6 ($p < .001$)	8 ($p < .001$)

for "appropriate" behavior were so obvious as to limit the nature of the responses.

Results

Of 22 predicted relationships between the Rorschach test and the TAT (Table 2), six were significant at the 5 cent level of confidence and twelve at the 10 per cent level. Marked contrast occurred, however, when affect variables based on the TAT themes were isolated, as five were significant at the 5 per cent level and nine were significant at the 10 per cent level. Comparison of Rorschach and TAT-overt variables reveals that of eight predicted relationships, only one was significant at the 5 per cent level of confidence and three at the 10 per cent level. The greater part of the consistency between the Rorschach test and the TAT appears to reside in Rorschach and TAT -theme inferences.

Of 22 predicted relationships between the Rorschach test and SCT, only one was significant at the 5 per cent level and five at the 10 per cent level. Relatively less consistency is manifested between these two tests than between any other two tests.

Of 27 predicted relationships between the TAT and the SCT, nine were significant at the 5 per cent level of confidence and twelve at the 10 per cent level. Of 13 predicted relationships between the SCT and TAT themes, three were significant at the 5 per cent level and four significant at the 10 per cent level. Of 14 predicted relationships between SCT and TAT-overt variables, six were significant at the 5 per cent level and eight at the 10 per cent level. The greater part of the consistency between the SCT and the TAT appears to reside in SCT and TAT-overt inferences.

To make the results between tests roughly comparable, the number of significant relationships was tested by the binomial expansion [28] to determine if that number was sufficiently great to indicate a nonchance occurrence. However, any interpretation of the resulting p values must necessarily be made with caution because of the lack of independence of the measures, the assumption on which such a test is essentially based.

Discussion

The extent of the relationships generally supports the conclusion that there is consistency in an individual's responses to tests of varying degrees of ambiguity.[2] That this consistency is found not in the overt manifestations of behavioral responses, but rather in an inferred explanatory construct related to motivational aspects of behavior, would seem consistent with the classic study by Frenkel-Brunswik [13].

It is obvious that the demonstrated consistency is far from perfect. To what degree this is dependent upon the somewhat arbitrary categorizing of test stimuli (particularly in regard to the TAT), on the inadequacy of the scoring system, or on factors more directly related to the subjects is not known.

The individuals chosen for this study comprised outpatients at a mental hygiene clinic. It seems likely that, as recently suggested by Hanfmann and Getzels [14], the degree of consistency between expressions elicited from different levels may be a function of personality integration. Greater consistency could then have been expected with a group of "normal" subjects, and less consistency in a group of frank psychotics. Only further research could verify this assumption, however.

In spite of the lack of perfect consistency, one point seems unequivocal: There is no support for the suggestion sometimes made that the opposite, or converse, of a given affect will appear at more overt levels as contrasted with that which exists at a deeper level. The trend is definitely in the direction of consistency rather than that of antithetical or diametrically opposed relationships, i. e., reaction formations.

[2]An analysis involving the intra-individual consistency *within* tests as reflected by the generalization of a given affect from mother to females, people, and authority figures, as well as from father to males, people, and authority figures, had revealed that the generalization within the various tests was only slightly above that expected on a chance basis. This finding would appear consistent with McClelland's statement that "while a set of generalized expectations about fathers and mothers or parent-child relationships probably do develop . . . they almost certainly do not *derive in any simple fashion* from the actual characteristics of the parents or the child's relationships to the parents . . . . There are many influences which enter into the formulation of such generalized conceptions" [20, p. 308].

This generalization, however, would hold only for whatever levels of functioning are represented by the chosen tests. This is an important stipulation as preliminary research had demonstrated no significant agreement between inferred attitudes to mother and father taken from SCT data and feelings expressed toward either parent on a rating scale. While apparently disqualifying SCT data as necessarily representative of overt manifestations, in the sense of being inwardly accessible and chosen for outward reporting, the finding would not be incompatible with the clinical impression that the SCT is a projective technique in the real sense of the word.

It is interesting to note that least consistency was revealed between the Rorschach and the SCT, tests representing the extremes in degree of stimulus ambiguity, while greatest consistency was found between each of these tests and the TAT. The findings perhaps could have been predicted on the basis of perceptual theory. Bruner has written: "Generally speaking, the less the structuredness or univocality of the stimulus input, the more striking the role of directive nonsensory factors will be in determining the perceptual organization" [6, p. 161].

Also related to the nature of the demonstrated relationships is the fact that the levels of inference involved, as well as the levels of personality function assumed to be tapped by the tests, are roughly equivalent, suggesting that these results are not fortuitous.

Thus, in spite of research evidence which would suggest that behavioral measures are inconsistent and unpredictable, it would appear that lawful relationships can be found in behavioral responses and that responses superficially dissimilar may nevertheless reflect a basic consistency and orderliness in perception and personality. That this can be demonstrated in the case of quantifiable variables which have meaning in terms of both projective and personality theory is evidence that a dynamic formulation of personality does not preclude the possibility of meaningful quantitative analyses of clinical data.

Summary

The intra-individual consistency of the inferred affect of responses to projective tests of varying degrees of ambiguity was studied. Variables were established in terms of given affect categories related to specified interpersonal figures. These categories were dichotomized about a median, and the statistical method used to determine the relationships between them was chi square.

Analysis of data demonstrated greatest consistency between the Rorschach test and the TAT, and between the TAT and the SCT. Least relative consistency was present between the Rorschach and SCT. The greater part of the consistency of the TAT with the Rorschach test lay in ratings based on TAT themes, while the greater part of the TAT's consistency with the SCT resided in ratings based on the overtly attributed characteristics of the varying interpersonal figures. The greatest consistency appeared among variables based on levels of inference as well as on those levels of personality functioning which approximate each other. While the obtained relationships showed far from perfect consistency, the extent of the demonstrated relationships justified the conclusion that there is a basic consistency and orderliness in perception and personality which may underlie the variability in behavioral responses.

Received December 1, 1953.

References

1. Abt, L. E., & Bellak, L. *Projective psychology.* New York: Knopf, 1950.
2. Beck, S. J. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune & Stratton, 1949.
3. Bender, I. E., & Hastorf, A. H. The perception of persons: forecasting another person's responses on three personality scales. *J. abnorm. soc. Psychol.,* 1950, **45,** 556-561.
4. Blake, R., & Wilson, G. Perceptual selectivity in Rorschach determinants as a function of depressive tendencies. *J. abnorm. soc. Psychol.,* 1950, **45,** 459-472.
5. Bray, D. W. The prediction of behavior from two attitude scales. *J. abnorm. soc. Psychol.,* 1950, **45,** 64-84.
6. Bruner, J. S. Perceptual theory and the Rorschach test. *J. Pers.,* 1948, **17,** 157-168.
7. Carp, Frances M. Psychological constriction in several projective tests. *J. consult. Psychol.,* 1950, **14,** 268-275.
8. Cronbach, L. J. Statistical methods applied to Rorschach scores: a review. *Psychol. Bull.,* 1949, **46,** 393-429.
9. DeVos, G. A quantitative approach to affect-

ive symbolism in Rorschach responses. *J. proj. Tech.*, 1952, **16**, 133-150.
10. Elizur, A. Content analysis of the Rorschach with regard to anxiety and hostility. *Rorschach Res. Exch.*, 1949, **13**, 247-284.
11. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).
12. Forer, B. R. A structured sentence completion test. *J. proj. Tech.*, 1950, **14**, 15-30.
13. Frenkel-Brunswik, Else. Motivation and behavior. *Genet. Psychol. Monogr.*, 1942, **26**, 121-265.
14. Hanfmann, Eugenia, & Getzels, J. W. Studies of the sentence completion test. *J. proj. Tech.*, 1953, **17**, 280-294.
15. Jones, L. V. Tests of hypotheses: one-sided alternatives. *Psychol. Bull.*, 1952, **49**, 43-46.
16. Kelly, E. L., & Fiske, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univer. of Michigan Press, 1951.
17. Korner, Anneliese. Theoretical considerations concerning the scope and limitations of projective techniques. *J. abnorm. soc. Psychol.*, 1950, **45**, 619-627.
18. Lazarus, R. S. Ambiguity and nonambiguity in projective testing. *J. abnorm. soc. Psychol.*, 1953, **48**, 443-445.
19. Lewis, D., & Burke, C. J. The use and misuse of the chi-square test. *Psychol. Bull.*, 1949, **46**, 433-489.
20. McClelland, D. C. *Personality*, New York: William Sloane Associates, 1951.
21. Meer, B., & Singer, J. A note on the father and mother cards in the Rorschach inkblots. *J. consult. Psychol.*, 1950, **14**, 482-484.
22. Orange, A. J. Consistency of manner of apperception in tasks of varying structure. Unpublished doctor's dissertation, Univer. of Chicago, 1951.
23. Rosen, E. Symbolic meanings in the Rorschach cards: a statistical study. *J. clin. Psychol.*, 1951, **7**, 239-244.
24. Rosenzweig, S. Idiodynamics in personality theory with special reference to projective techniques. *Psychol. Rev.*, 1951, **58**, 213-223.
25. Samuels, H. The validity of personality-trait ratings based on projective techniques. *Psychol. Monogr.*, 1952, **66**, No. 5 (Whole No. 337).
26. Sanford, R. N., *et al.* Physique, personality, and scholarship. *Monogr. soc. Res. Child Develpm.*, 1943, No. 34.
27. Tupes, E. C. An evaluation of personality trait ratings obtained by unstructured assessment interviews. *Psychol. Monogr.*, 1950, **64**, No. 11 (Whole No. 317).
28. Wilkinson, B. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, **48**, 156-158.
29. Wittenborn, J. R. The implications of certain assumptions in the use of the Thematic Apperception Test. *J. consult. Psychol.*, 1950, **14**, 216-225.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Rorschach Concepts of Normality[1]

Ann Lawler Brockway, Goldine C. Gleser, and George A. Ulett

Washington University School of Medicine

St. Louis, Missouri

During the past decade the Rorschach test has been administered more and more frequently to "normal" individuals, either as military personnel or as members of control samples in experimental research. As a result of the increasing accumulation of data on such subjects, many of the widely held concepts regarding what constitutes a "normal" Rorschach protocol have failed to be substantiated. Papers by Linn [9], Brussel and Hitch [3], and Wischner, Rotter, and Gillman [12] have presented summaries of Rorschach findings among military personnnel, while Beck *et al.* [1] and Cass and McReynolds [4] have published data on samples of normal adults from varied occupations. Contributions such as these should cause, and to some degree have caused, clinicians to be more cautious in interpreting certain Rorschach characteristics as signs of pathology, and have made it necessary to revise many of the widely held tenets of what constitutes a normal protocol. It is becoming more evident that the perception of inkblots is affected by factors such as sex, race, age, intelligence, education, and set and that individual differences cannot be interpreted on the basis of personality variables alone. It appears necessary, therefore, that a careful accumulation of Rorschach data on representative samples of the total population stratified with regard to these factors be obtained, so that interpretations of pathology may be made on the basis of appropriate norms for each individual.

The point has been raised, however, particularly by Klopfer [8] and more recently by Beck *et al.* [1] that normality when regarded as a statistical average may differ widely from a qualitative definition of normality as meaning well balanced. This is admittedly true, but it does not mean that we can make interpretations which are not confirmed by the empirical data. A possible solution to this dilemma is to base norms on individuals who are not only engaged in everyday activities implying adequate adjustment but who have also been interviewed by a psychiatrist and considered to be relatively free from neurotic symptoms. Such a criterion, while far from perfect, would tend to yield a more homogeneous sample with regard to personality adjustment. The purpose of this paper is to make a beginning in this direction by summarizing Rorschach scores of a group of white subjects homogeneous with regard to sex, age, education, and set, and psychiatrically screened on the basis of personality adjustment. The scores for the well-adjusted members of this group are compared with those of the maladjusted, and also with scores obtained on a group of psychiatric patients, in order to examine the influence of personality adjustment on Rorschach percepts. Our results add further evidence that many of the rule-of-thumb "signs of normality" used by interpreters of the Rorschach are inadequate.

Procedure

A group of 151 men without history of psychiatric disease, and 40 male patients with a variety of psychiatric disorders in which manifest anxiety was considered the predominant symptom, were administered the Rorschach test by one female examiner and screened as to psychological adjustment by one male psychi-

[1]This report was completed under Air Force Contract Number 33(038)-13884. It is one of a series under R and D Project 21-37-002, Division of Psychiatric Screening of Flying Personnel, Department of Clinical Psychology, United States Air Force School of Aviation Medicine.

Table 1

Statistics on Age and Completed Education for 151 Normal Subjects and 40 Patients

Group	No. of cases	Age (years)		Median	Education completed (grade)			
		Median	Range		Distribution			
					8–11	12	13–15	16 and over
Adjusted	126	22	17–36	14	6	31	38	51
Maladjusted	25	22	18–33	15	0	7	6	12
Total normals	151	22	17–36	14	6	38	44	63
Patients	40	29	18–36	12	9	11	9	11

atrist. The Rorschach was administered second in a battery of psychological tests. The detailed procedure has been reported elsewhere [11]. All of the normal group were paid volunteers, oriented to the effect that the tests administered were being investigated as part of a proposed battery to be used in screening Air Force personnel. In this way it was felt that the set toward the test was relatively constant. The patients had been informed by their psychiatrists that the results of the tests would provide information useful in therapy. Thus their set differed somewhat from that of the "normals." All Rorschach protocols were scored as to determinants and content according to the Klopfer and Kelley system [8] and for location, form level, populars, and originals according to Hertz [6]. Additionals were counted as one-half in the tabulation of determinants and were given full value in counting the number of space responses. They were not included in the remaining scores.

Table 1 presents vital statistics (age and education) on the sample classified according to psychiatric appraisal. It may be noted that the patients tend to be slightly older than the normal subjects and include a somewhat smaller proportion of college graduates. The subjects for the normal group consisted of 38 students from a large metropolitan university, 39 students of a Protestant seminary, 36 members of the Air Force reserves, 15 regular Army personnel, 20 new Army recruits, and 3 miscellaneous. The individuals were engaged in occupations ranging from unskilled laborer to semiprofessional, but the majority of the subjects were either "white collar workers" or students. The patients were obtained from private practice or clinics, were for the most part gainfully employed, and also included some college students.

Results

Table 2 presents the 50th percentile scores in the major Rorschach scoring areas for the total normal group and the 50th percentile ranks found by Cass and McReynolds [4] on a group of 104 "normals" (median age, 34; median grades completed, 12). Their figures are shown only for those categories for which the scoring methods were the same. In comparing these values it should be noted that the statistics used by Cass and McReynolds are not quite identical with the exact median point used in this study. Even so, their findings are

Table 2

Median Scores for Various Rorschach Scoring Categories in a Normal Sample of Young Males Compared to 50th Percentile Ranks Obtained by Cass and McReynolds on Comparable Scoring Categories

Rorschach category	Present sample $N = 151$	Cass and McReynolds $N = 104$
R	26.2	22.5
M	2.0	2.0
FM	3.7	3.5
m (*m*, *mF*, *Fm*)	1.2	.5
FK	0.2	0.0
K (*K*, *KF*)	0.1	0.0
k (*k*, *kF*)	0.2	0.0
F	11.1	7.5
Fc	1.4	2.0
c (*c*, *cF*)	0.7	0.0
C′ (*c′*, *C′F*, *FC′*)	0.9	1.0
FC	0.9	1.5
CF	2.4	2.0
C (*C*, *Cn*, *C sym.*)	0.0	0.0
Σ*C*	2.8	3.0
Content		
H	2.4	2.5
Hd	1.0	0.5
A	9.1	9.0
Ad	1.4	1.0
P	5.6	5.5

very similar to those presented here, which is all the more remarkable since their sample included 46 females and a somewhat wider age and educational range than did the sample in our study. The major difference appears to be that they obtained a smaller median number of responses with an equivalent reduction in the median *F* response. They also obtained somewhat fewer responses determined by *c* and *m* and more categorized *FC*. It would be interesting to determine whether these differences could have been due to the inclusion of females in Cass's study. In both Cass's data and ours it may be noted that *FM* is greater than *M* and *CF* is greater than *FC*, although the reverse inequalities have usually been accepted as indicative of good normal adjustment.

Beck's results are not directly comparable to ours because of differences in scoring categories. Also, he presents his results in terms of means and standard deviations, despite the ex-

Table 3

Medians, 20th, and 80th Percentile Points on Various Rorschach Scoring Categories for a Normal Sample Classified as to Personality Adjustment and for a Sample of Psychiatric Patients

Rorschach Category	Adjusted normals ($N = 126$) Percentile scores			Maladjusted normals ($N = 25$) Percentile scores			Psychiatric patients ($N = 40$) Percentile scores		
	20th	50th	80th	20th	50th	80th	20th	50th	80th
R	15.9	26.6	39.9	14.7	24.9	30.5	15.5	22.2	33.8
Location									
W (W, W, DW)	4.7	8.4	13.6	6.2	9.3	13.5	4.0	8.0	13.8
D	5.8	12.1	22.3	4.3	11.0	18.5	3.5	10.5	19.5
dr	.8	2.9	7.5	.5	1.9	4.5	0	1.8	7.1
S	.9	2.9	5.9	.5	2.4	4.2	1.4	2.9	5.8
W%	16.7	33.2	58.7	21.8	41.8	70.5	16.3	30.5	65.5
D%	27.1	48.4	62.2	20.5	46.0	61.8	20.5	45.5	64.5
dr%	3.6	12.1	22.7	3.7	8.8	14.5	0	8.2	28.0
S%	3.6	10.1	18.8	.2	9.5	14.5	6.3	11.8	25.5
Determinants									
M	.9	2.0	3.4	.2	1.2	2.8	.2	1.1	3.1
FM	1.3	3.9	6.9	1.8	2.8	5.8	1.2	3.2	4.8
m (m, mF, Fm)	.1	1.0	2.7	.1	2.0	3.2	0	.9	1.8
FK	0	.2	1.2	0	.2	1.2	0	.1	1.0
K (K, KF)	0	.1	1.0	0	.1	1.0	0	.1	.9
k (k, kF)	0	.2	1.2	0	.4	1.2	0	.8	1.6
F	6.1	11.5	20.7	4.4	9.6	16.8	6.2	10.5	17.3
Fc	.2	1.4	3.1	.2	1.2	3.0	.1	1.0	2.8
c (c, cF)	0	.6	1.9	0	1.0	3.8	0	.2	1.6
C' (C', C'F, FC')	0	.8	2.2	.2	1.2	2.2	.1	.9	2.1
FC	.1	1.0	2.2	0	.7	2.2	0	.8	1.8
CF	.4	2.2	4.2	1.9	3.2	5.2	.6	1.6	3.4
C (C, Cn, Csym.)	0	0	.2	0	.1	.2	0	.1	.25
ΣC	.9	2.6	4.6	1.8	3.9	5.8	.1	1.9	4.4
Content									
H	.9	2.6	4.5	.9	1.9	4.0	.6	1.5	3.1
Hd	0	1.1	3.6	0	.6	2.5	0	1.6	4.0
A	5.4	9.4	13.5	5.0	8.0	11.0	3.0	6.7	9.8
Ad	.2	1.9	2.0	0	.7	3.5	0	.3	3.5
At	.6	2.3	4.5	1.2	2.2	3.9	.3	2.1	5.8
Sex	0	0	.3	0	0	.3	0	.1	.4
Clouds	0	.1	.8	0	.1	1.0	0	.1	.2
Blood	0	0	.4	0	.2	.7	0	.1	.4
Ob., Art, Arch.	.4	2.4	5.7	.5	2.6	4.1	.7	2.3	4.5
Nat., Geo.	.8	3.3	6.5	1.2	3.4	8.8	.7	3.0	7.0
P	4.1	5.7	8.2	3.3	5.1	7.0	3.3	5.5	7.2
O	1.6	5.0	11.0	3.5	5.0	9.5	1.3	3.8	9.1

treme skewness of distribution of the scores in most categories. However, we might note that he also found mean *CF* greater than mean *FC* for his sample.

Table 3 presents the 20th, 50th, and 80th percentile scores in the major Rorschach scoring areas for the screened normals, the maladjusted normals, and the patients. It is believed that the "adjusted" group is a fairly representative sample of young adult white males of higher than average educational level and relatively good personality adjustment. Therefore, the middle 60% range of scores should approximate that which may be considered usual or typical for such a group.

From Table 3 it can be noted that the adjusted normals are slightly more productive of responses than either the maladjusted normals or the patients. Some interesting points of contrast between the adjusted and maladjusted normals of this sample are the latter's greater emphasis on whole responses at the expense of other location areas; their greater use of *c*, *m*, *C'*, and *CF* as determinants; and *blood*, *nature*, and *geography* as content categories. The adjusted normals on the other hand apparently use more *M*, *H*, and *A* than do the maladjusted subjects. Of all these categories however, only *M* and *m* yield a statistically significant difference at the optimal cutoff point.

The patients differ from the adjusted normals in their relatively greater emphasis on *S* responses. They also make greater use of *k*, *C*, and *C'* as determinants and *Hd* and *sex* as content particularly when these scores are considered relative to the total number of responses. On the other hand, it is particularly interesting to note that in some categories where differences are generally hypothesized, no differences are found for this sample. Thus, the adjusted controls utilized *K* (usually interpreted as free-floating anxiety) and *m* (interpreted as indicating tension and repression) to as great an extent as did the anxious patients.

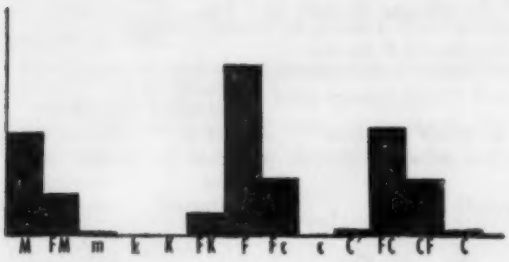
Although trends were not necessarily to be expected in going from the adjusted normals through the maladjusted normals to the patients, since the maladjusted normals were so rated on a wide variety of symptoms whereas the patients were chosen primarily on the basis that they had manifest symptoms of anxiety, nevertheless some trends are apparent. It appears that as degree of maladjustment increases there is a diminution in productivity and also in the use of *M* as a determinant and *H* and *A* as content. On the other hand, there is an increase in the use of *k* and pure *C* as determinants. Unpublished research done at this laboratory has demonstrated that when *FC'*[2] is omitted from the *C'* category and cutting points are established with respect to number of *R*, the incidence of this determinant (*C'* and *C'F*) in protocols can be used to discriminate adjusted from the maladjusted subjects and the patients (*C'* + *C'F* less for adjusted subjects).

One of the most important facts to be obtained from Table 3 is that the range of scores for the middle 60% of the adjusted normals is often greater than is generally considered optimal by Rorschach workers (Klopfer, Brussel, and Hitch). This is particularly true of *W*% (17% – 59%), *dr*% (4% – 23%), *S*% (4% – 19%), *FM* (1.3 – 6.9), and *CF* (0.4 – 4.2). For other categories the middle range is shifted somewhat lower or higher than is postulated. It is suggested that optimal norms for Rorschach categories might better be developed by obtaining data on samples such as the present, in which psychiatric ratings are utilized to screen out the obviously maladjusted and then the middle 60% used to define the expected or optimal range.

Figure 1 presents the psychogram (using median scores) of the psychiatrically screened normals studied in this investigation along with a textbook illustration of the "normal" Rorschach psychogram [2, p. 128]. A comparison of these psychograms reveals that the screened "normals" used *F*, *FM*, *m*, *k*, *K*, *c*, and *C'* to a relatively greater extent than the hypothetical normal, with *M*, *FK*, *Fc*, *FC*, and *CF* being used to a lesser degree. It will also be observed that the ratios of *M* to *FM* and *FC* to *CF* differ in direction in the two psychograms, with *FM* and *CF* exceeding *M* and *FC* respectively in the screened normal group.

Table 4 shows the incidence of Davidson's [5] Rorschach signs of adjustment in the psychiatrically classified subgroups. Using com-

[2]Use of *FC'* as a determinant was demonstrated to be a sign of good adjustment.



HYPOTHETICAL NORMAL PSYCHOGRAM

ADJUSTED NORMAL SAMPLE N=126

Fig. 1. Comparison of the hypothetical normal psychogram with that obtained on the present adjusted normal sample.

monly accepted Rorschach signs of adjustment but no outside criterion, Davidson reported the biserial correlation between the presence or absence of each sign and the total number of signs, thus relying on internal consistency as the test of validity. By comparing the percentage of each group showing a sign, it may be seen that a number of the signs discriminate in the direction opposite to that predicted by Davidson. Thus a larger percentage of the patients had $M \geq FM$ and $Dr+S \leq 10\%$ than did either group within the normal sample.[8] There was also a reversal in expected direction in the case of *Sum C*> $(Fc+c+C')$; $F\% \geq 20$ and ≤ 50; $\%R$ *VIII-X* ≥ 40 and ≤ 60; $A\% \leq 50$ and refusals, in that the percentage of maladjusted normals, and in some cases of patients, showing the signs was greater than of the adjusted. The signs that held up, though not always to a statistically significant degree even under a one-tailed hypothesis, were $M \geq 3$; $P \geq 4$ but $< .30\ R$; $R > 20$; $FC \geq 2$; no pure C; $FK + Fc \geq 2$; and $FC \geq CF$. For the latter sign it should be noted that it is necessary to exclude cases for which $FC = CF = 0$ in order to obtain the discrimination. Davidson did not specify this exception. Also for the P scores it was necessary to alter the criterion where R was less than or equal to 13. For these cases 4 P's were considered a sign of good adjustment.

[8]The symbol $\geq$ means greater than or equal, and the symbol $\leq$ means less than or equal.

Special note should be made of the incidence of shading and color shock in the three psychiatrically classified groups. Using the criteria proposed by Munroe [10], we found that only three of the subjects in the total normal group were *without* evidence of color shock (1.6% of the adjusted normals and 4% of the maladjusted normals) and 14 were *without* evidence of shading shock (7.9 and 16%). Though 100% of the patients showed some evidence of color shock, 10% of this group were without evidence of shading shock as compared to 7.9% of the adjusted normals. These findings are remarkable not so much as indicating a reversal in expected trend but as pointing out that what has been termed "shock" actually occurs to some extent in almost all Rorschach protocols. Thus these signs, at least when defined according to Munroe's criteria, cannot be used in detecting maladjustment. This criticism has also been posed in an article by Keehn [7], who makes a review of the literature with reference to color shock and concludes that the sign is of questionable value in the diagnosis of neuroticism.

That as great a proportion of the adjusted normals refused one or more cards as did the maladjusted is rather surprising, since card refusal has been accepted by most clinicians as pathological. The highest incidence of card rejection in our normal sample occurred in a group of 15 career military men, 13 of whom were sergeants and 2 commissioned officers. This subgroup had 5 members (33 1/3%) who rejected one or more cards. This finding is in agreement with the report of Linn [9] that "well-adjusted" military personnel frequently reject Rorschach cards. In our sample it was noted in particular that Cards V and/or VI were rejected more frequently by the "ad-

Table 4

Application of Davidson's Signs of Adjustment on Rorschach to Present Sample Classified According to Psychiatric Criterion

Rorschach sign of adjustment	Percentage of group in which sign was present			Probability of no difference (One-tailed hypothesis)	
	Adjusted normals	Maladjusted normals	Patients	Adjusted normals versus maladjusted normals	Adjusted normals versus patients
$M \geq FM \neq 0$	27.8	16.0	30.0	.11	Reversed†
$M \geq 3$	40.5	20.0	22.5	.03*	.02*
$\Sigma C > Fc + c + C'$	52.4	64.0	45.0	Reversed	.21
$20 \leq F\% \leq 50$	59.5	64.0	57.5	Reversed	—
$Dr + S \leq 10\%$	37.3	48.0	55.0	Reversed	Reversed
$P \geq 4 < .30\ R$ (Except when $R \leq 13$, $P = 4$)	69.0	60.0	57.5	.19	.09
$R > 20$	68.3	64.0	55.0	—	.06
$FC \geq CF \neq 0$	36.5	20.0	30.0	.06	.22
$FC \geq 2$	31.0	24.0	20.0	.24	.09
No pure C	86.5	76.0	80.0	.09	.16
$\%\ R$ VIII-X $40 \leq R\% \leq 60$	34.1	36.0	20.0	Reversed	.05*
$FK + FC \geq 2$	62.7	60.0	47.5	—	.04*
$W{:}M > 1{:}1$ and $< 3{:}1$	21.4	12.0	20.0	.15	—
$A\% \leq 50\%$	67.5	80.0	72.5	Reversed	Reversed
No color shock†	1.6	4.0	0	Reversed	—
No shading shock‡	7.9	16.0	10.0	Reversed	Reversed
No refusals	87.3	88.0	77.5	Reversed	.07

* Statistically significant.
† By reversed is meant that the adjusted normals have the lower percentage.
‡ As defined by Munroe [10].

justed normals" than by maladjusted normals or patients.

Summary

Selected percentile scores were presented for the major scoring categories of Rorschach protocols of 151 white males, age 17-36, all of whom had had at least an eighth grade education. All Rorschach data were obtained and scored by one psychologist, and the testing was done under uniform test conditions and set.

These median scores for the total "normal" group were compared with the 50th percentile ranks obtained by Cass and McReynolds on comparable scoring categories. Our sample was found to have a somewhat higher median number of responses, which was reflected mainly in the *F* category, but for the most part the obtained scores were very similar, giving encouragement to the idea that stable and meaningful norms can be obtained for the Rorschach. Our sample confirmed their finding that *FM* is typically greater than *M* and *CF* greater than *FC* in the normal protocol. The latter inequality is also confirmed by Beck's data.

Since most Rorschach interpretations are based on the concept of normal in the sense of well adjusted, rather than average, it was felt that the most meaningful norms to be obtained from the present sample would be those based on screened normals. All subjects had been interviewed by a psychiatrist and rated as to over-all adjustment. The psychiatric rating was used to divide the sample into adjusted and maladjusted normals, and the 20th 50th, and 80th percentile scores were presented for these subgroups. In addition, the scores of 40 patients with symptoms of manifest anxiety were presented for comparison. While some commonly accepted signs tended to differentiate the maladjusted and/or the patients from the screened normals, many others did not hold for this sample. In particular, the adjusted group used *K* and *m* to as great an extent as did the anxious patients. The middle 60% of the group showed a wider range of

W%, *dr%*, *S%*, and *FM* than is usually considered optimal.

Color shock and shading shock as defined by Munroe were found to some extent in all but a very few protocols and thus, as presently defined, are of no value in detecting maladjustment.

It appears that a great deal more empirical research is needed to determine meaningful norms for the Rorschach. In particular, the influence of such factors as sex, race, age, education, and socioeconomic background must be determined, and separate norms obtained when necessary in order to take these factors into account. Also, since Rorschach interpretation is, in general, based on the concept of the optimal or well-adjusted Rorschach protocol, it appears reasonable to obtain norms for screened normal groups. The present paper makes a contribution in this direction by indicating the median score and middle 60% range for a group of young, psychiatrically screened, normal, white males of more than average education. It is hoped that comparable data will be published for other relatively homogeneous groups. Such norms would greatly increase the efficiency and validity of clinical interpretation.

Received November 20, 1953.

References

1. Beck, S. J., Rabin, A. J., Thiesen, W. G., Molish, H., & Thetford, W. N. The normal personality as projected in the Rorschach test. *J. Psychol.*, 1950, **30**, 241-298.
2. Bell, J. E. *Projective techniques.* New York: Longmans, Green, 1948.
3. Brussel, J. A., & Hitch, K. S. *An introduction to Rorschach psychodiagnostics in military and civilian psychiatry.* Utica, New York: State Hospitals Press, 1947.
4. Cass, W. A., Jr., & McReynolds, P. A contribution to Rorschach norms. *J. consult. Psychol.*, 1951, **15**, 178-184.
5. Davidson, Helen H. A measure of adjustment obtained from the Rorschach protocol. *J. proj. Tech.*, 1950, **14**, 31-38.
6. Hertz, Marguerite R. *Frequency tables for scoring responses to the inkblot test.* (3rd Ed.) Cleveland, Ohio: Western Reserve Univer. Press, 1951.
7. Keehn, J. D. Rorschach validation. II. The validity of colour shock in the diagnosis of neuroticism. *J. ment. Sci.*, 1953, **99**, 224-234.
8. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1946.
9. Linn, L. The Rorschach test in the evaluation of military personnel. *Rorschach Res. Exch.*, 1946, **10**, 20-27.
10. Munroe, Ruth. The inspection technique: a method of rapid evaluation of the Rorschach protocol. *Rorschach Res. Exch.*, 1944, **8**, 46-70.
11. Ulett, G. A., Gleser, Goldine, Lawler, Ann, & Winokur, G. An experimental investigation of development of an EEG index of anxiety tolerance by means of photic stimulation — its validation by psychological and psychiatric criteria. *USAF Sch. Aviat. Med. Proj. Rep.*, 1952, Proj. No. 21-37-002, Rep. No. 4.
12. Wischner, G. J., Rotter, J. B., & Gillman, R. D. Projective techniques. In S. W. Bijou (Ed.), *The psychological program in AAF convalescent hospitals.* Washington, D.C.: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 15.) Pp. 144-157.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

An Evaluation of the Memory-for-Designs Test[1]

Alvin R. Howard
VA Hospital, Chillicothe, Ohio[2]
and Donald J. Shoemaker
Ohio State University

Attempts at psychometric isolation of patients with the diagnosis "organic brain damage" have been noticeably inconclusive. One of the techniques offered, but to date little reported on, is Graham and Kendall's Memory-for-Designs Test. [1, 2].

Forty-two male, neuropsychiatric patients, evenly divided between experimentals and controls, comprised the sample studied. The experimental group was composed of patients diagnosed both as "psychotic" and "organic"; the control group consisted of persons diagnosed as "psychotic" only. In selecting subjects, the criteria of the test authors were employed. Individual matching was accomplished whereever possible. Compared with the male and female *Ss* employed by Graham and Kendall, the present sample is older, completed fewer grades of formal schooling, and does not contain as wide a variety of diagnoses.

Independent scoring of individual designs (on training protocols) by the authors resulted in a 93% agreement, lending favorable support to Graham and Kendall's correlation of .99 between independent ratings of total scores.

Mean visual-motor raw and difference scores differentiated experimental and control groups beyond the 5% level, also supporting earlier findings of the test authors, although to a lesser degree. Similarly supportive were various correlations duplicated by the present investigators. The Pearson *r*'s between visual-motor score and age, education, and vocabulary, respectively, were .26, —.31, and —.31. The multiple *r* between visual-motor score and age, vocabulary, and grade was .44; between visual-motor score and age and vocabulary was .42. The corresponding coefficients reported by Graham and Kendall were comparably low. Less supportive results occurred in relation to the distribution of visual-motor scores. In respect to both raw and difference scores, a much larger percentage of our control group subjects fell above the critical scores denoting brain damage. Whereas only 4% of Graham and Kendall's "normative adults" had difference scores assigning them to the brain-damage category, our corresponding figure was 19%; similar low percentages were noted by them on other control groups when raw scores were used, whereas 28% of our controls were selected by this method. Experimental group findings by both raw and difference methods are more nearly alike, our respective 52% and 43% approximating their 45%, 50% and 50%. Our results thus suggest the use of increased caution in attributing brain damage to individual neuropsychiatric patients scoring high by either the raw or difference method.

Brief Report
Received March 23, 1954.

[1] An extended report of this study may be obtained without charge from Alvin R. Howard, Veterans Administration Hospital, Chillicothe, Ohio, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4254 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[2] From the Veterans Administration Hospital Chillicothe, Ohio.

References

1. Graham, Frances K., & Kendall, Barbara S. Performance of brain-damaged cases on a memory-for-designs test. *J. abnorm. soc. Psychol.*, 1946, **41**, 303-314.
2. Kendall, Barbara S., & Graham, Frances K. Further standardization of the Memory-for-Designs Test on children and adults. *J. consult. Psychol.*, 1948, **12**, 349-354.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Determinant Shift of Psychoneurotic and Psychotic Patients[1]

Robert G. Gibby and Bernard A. Stotsky
VA Regional Office, Detroit, Michigan

In a previous paper [2], three areas of the relationship of the Rorschach free association and inquiry were investigated. These were: (*a*) their similarity; (*b*) the comparative effect of the stimulus value of the examiner on each; and (*c*) the relationship of the personality characteristics of the subject to changes from free association to inquiry.

Briefly summarized the findings were:

1. The free association record differs significantly from that of the inquiry.

2. The free association is less susceptible to examiner influence than the inquiry.

3. Personality characteristics, as measured by the free association of the Rorschach, of those individuals changing the most differ significantly from those changing the least.

These findings suggested that there were dimensions of the Rorschach worthy of investigation other than those presently tapped by the traditional scoring method. The changes in Rorschach determinants from free association to inquiry, which we shall call "Determinant Shift," especially offered promise. These changes might very well be related to the capacity of a subject to respond to new stimuli and to interact in a structured interpersonal relationship. For instance, the degree of shift of determinants has been used successfully to predict the ability of a subject to form a lasting therapeutic relationship [3].

The calculation of the Determinant Shift score is a relatively simple matter. First, the free association responses are scored independently of the material elicited on the inquiry. Next, the combined free association and inquiry responses are scored. These scorings produce two psychograms, one for free association and one for the inquiry. The score for each determinant on the free association is then subtracted from the score following inquiry. This results in a shift score for each determinant. The Determinant Shift score for all determinants except *F* will usually be either zero or positive, whereas for *F* the score will almost always be zero or negative since the shift is most often from *F* to some other determinant. Consequently, the shift in *F* may usually be regarded as reflecting the total amount of shift.

The administration of the Rorschach is made up of two interdependent, yet quite different, procedures. The first, the free association, is relatively unstructured. The subject is simply requested to free associate to the Rorschach cards as they are presented to him. The second, the inquiry, is more highly structured. Here the subject is asked to explain and to clarify the previously given associations, to delineate the area of the blot utilized and to describe what characteristics of the blot suggested the response. The inquiry itself requires greater attention to, and much more direct interaction with, the examiner. It may be hypothesized that the more sensitive the individual, the more responsive he will be to the examiner. We would then expect greater shift to be shown on determinants from free association to inquiry by more sensitive and more reactive subjects.

To test this hypothesis, a group of 40 psychoneurotic outpatients at a Veterans Administration clinic was compared with a group of 40 psychotic inpatients at a Veterans Administration general, medical, and surgical hospital. The specific prediction was that the

[1]From the VA Regional Office, Detroit, Michigan.

psychoneurotics would show greater shift on Rorschach determinants from free association to inquiry than would the psychotics. It was also predicted that they would show greater shift on variables supposedly more indicative of anxiety (*FY, Total Y*) than on those presumably related to the more stable aspects of the personality, such as color and movement.

The Rorschach determinants investigated were *F, M, FC, CF, C, FY, YF, Y, V, Total Y,* and *Total C.* Only those determinants were studied which yielded large enough frequencies to warrant analysis. Beck's scoring techniques were employed for scoring these determinants. Indices of determinant shift were then calculated for each record [1]. Reliability of the scoring was determined. For this purpose, 10 records previously scored by one of the authors were randomly selected from the entire sample and then rescored by a second person who had been trained previously in the scoring techniques used in the study. Pearson product-moment correlations between the two scores were computed for each determinant, both for the free association by itself and for the total response. For free association the mean reliability was .90; for total response the mean reliability was .94.

In order to control the variability possibly introduced by *R*, an attempt was made to match each psychotic record *exactly* with a psychoneurotic record for the total number of responses. Out of 240 available psychoneurotic records, perfect matches were possible for 37 of the 40 psychotics. For two of the remaining three, there was a difference in *R* of one response. For the other, the difference was three responses. The total difference of *R* for both groups was three. Each record was compared with its matched record for differences in determinant shift. Following this, two types of statistical analyses were performed. The nonparametric "sign" test was used. Here, for the 40 matched pairs of records, the number of times the shift in determinants from free association to inquiry in one group exceeded that of the other was compared for significance. Where the assumption of normality for the amount of shift was tenable, the *t* for correlated means was also utilized to compare differences on shift for each of the determinants.

In order to check the possibility that the shift from free association to inquiry might be a function of existing differences on free association, the two groups were compared on all determinants in the free association. No significant differences were found for any of the determinants.

Results

A previous study [2] demonstrated that the determinant shift for *F* may serve as an indicator of the shift in all determinants from free association to inquiry. This index was used in the present study to compare the total shift of the psychoneurotic with that of the psychotic group. The psychoneurotic groups showed, using the sign test,[2] more change, with the difference in the index of determinant shift for *F* being significant at the .001 level. The prediction that the psychoneurotic group would change more from free association to inquiry was thus upheld.

The indices of determinant shift for the shading variables (*FY* and *Total Y*) in each group were also compared by the sign test. The differences between the groups were significant for both variables. For *FY*, the difference was significant at the .01 level; for *Total Y*, at the .001 level.

The differences between the indices of determinant shift for the color and movement variables were not found to be statistically

[2]The first test of the hypothesis was by means of the sign test. A problem presented itself in terms of what to do with cases where the amount of shift from free association to inquiry was the same for both groups. The following procedure was finally adopted: The results of the sign test were plotted on a normal curve; first, going from left to right, the proportion of cases where the shift was greater for psychotics; then the ties; and finally the proportion of cases where the shift was greater for neurotics. Next, the distance between the z scores for the ties was found. Then, with the mean set at zero, half of the distribution of ties was plotted on one side and half on the other side of the mean. From that point, determining the theoretical frequencies of the ties was simply a matter of consulting the table of normal curve functions and adding the area from m to z on both sides of the mean. The proportion of the curve thus obtained was multiplied by 40 for the theoretical frequency of ties. The remaining area on each side of the curve was multiplied by 40 to give the theoretical frequency for the cases where psychotics showed more shift than neurotics and for cases where neurotics showed more shift than psychotics. The chi-square values (two degrees of freedom) between obtained and theoretical frequencies were then computed for the three signs.

significant. The index of shift for color tended to be higher, and that for movement lower, in the psychoneurotic group. The results of the comparison of the groups by determinants are summarized in Table 1.

Table 1

Comparison of Neurotics and Psychotics on "Determinant Shift" by the "Sign" Test

Determinant	P > N*	P = N**	N > P†	*p*
F	8	5	27	.001
M	12	21	7	.46
FY	7	8	25	.01
YF	3	31	6	.64
Y	—	39	1	—
V	3	31	6	.64
Total Y	5	9	26	.001
FC	7	16	17	.12
CF	7	21	12	.46
C	—	—	—	—
Total C	9	11	20	.11

* Instances in which psychotics shifted more than neurotics from free association to inquiry.
** Instances in which the amount of shift for neurotics and psychotics was equal from free association to inquiry.
† Instances in which neurotics shifted more than psychotics from free association to inquiry.

The mean differences in determinant shift for *F, FY, Total Y,* and *Total C* were then compared by means of the *t* test. The *t* test was not used for the other determinants inasmuch as the distributions were markedly skewed by the presence of many 0 scores. These results are summarized in Table 2.

Table 2

Comparison of Neurotics and Psychotics on "Determinant Shift" by the *t* Test

	Means				
Determinant	Psychotics	Neurotics	Diff.	*t*	*p*
F	−3.15	−5.15	2.00	2.78	.01
FY	1.23	2.38	1.15	3.19	.01
Total Y	1.42	2.93	1.51	3.37	.001
Total C	1.38	1.85	.47	1.15	.25

The findings for *t* agree very closely with those for the "sign" test. Neurotics show significantly greater shift for *F, FY,* and *Total Y*. They also show more shift for the color variables, but again, as for the sign test, not significantly.

From these findings it can readily be seen that the Determinant Shift scores for *F, FY,* and *Total Y* differentiate at a significant level of confidence between neurotics and psychotics in the predicted direction. Fifty-six out of 80, 70 per cent of the patients, are correctly identified when the median for the shift in *Total Y* for both groups is used as a point of division, with scores above the median classified as neurotics and scores below the median as psychotics. The number of cases correctly classified is significantly better than chance. Their greater ability to shift suggests that psychoneurotics show more sensitivity to the testing situation and respond to the additional stimuli of the inquiry more actively than do psychotics. The greater increase in shading from free association to inquiry among psychoneurotics may be indicative of the presence of more anxiety in their reactions than in those of the psychotics.

Among the determinants of the Rorschach, the amount of shift may well be the important variable differentiating psychotic and psychoneurotic patients. The present-day method of scoring the total response, including both free association and inquiry, tends to obscure such shifts. Examination of our data demonstrates the point clearly. Comparison of the two groups by the traditional method of scoring determinants shows significant differences only on the shading variables *FY* and *Total Y*. On the other variables the shift from free association to inquiry leads to a convergence of the groups. For example, neurotics show more *F* on free association than psychotics. As a result of the shift from free association to inquiry, psychotics end up with more *F*. Similarly, psychotics show slightly more *FC, CF, C, V, YF, Y,* and *M* on the free association; as a result of the shift, psychoneurotics end up with slightly higher scores for these variables. The differences which exist between the two groups in the amount of shift, however, tend to be concealed; in the case of *F,* a very significant difference.

Comparison of psychotics with most and least shift. In the previous study, in comparing the 40 psychoneurotics who shifted most with the 40 who shifted least, significant differences were found for *F, M,* and *FC* on free association. The tentative conclusion drawn from the data was that the group shifting least was more comfortable in the interpersonal situation and less ready to interpret the situation as

one of threat. The question arose as to whether the same held true for psychotics. Would a group of psychotics who shifted little compared with a group who shifted considerably show the same differences on *M*, *F*, and *FC*?

To test this hypothesis, 16 psychotics giving shift scores of 0 and 1 were compared with 15 giving shift scores of 5, 6, and 7. Since there was a significant difference for *R* between the two groups, the arc-sine transformation was used to control for *R*. Fisher's exact test was then applied to the data, the fourfold categorization being in terms of presence or absence of the determinant on free association. Significant differences at the five per cent level of confidence were found only for *M*, *CF*, and *C*, with the group that showed the most shift giving more *M* and *CF* responses, and the group that showed the least shift giving more pure *C* responses.

The obvious discrepancy between the findings for psychoneurotics and those for psychotics does not appear to be due to chance in view of the significance of the differences found for the two groups. It may be that the meaning of shift for psychotics is different than for psychoneurotics. Among psychotics, shift may be indicative of retention of better emotional and intellectual control. Patients who do not shift may have become unable to respond effectively to the situation. Among neurotics, on the other hand, shift may be indicative of some loss of control, while lack of shift may be characteristic of a person with brittle, rigid, yet intact, defenses. The use of external criteria of adjustment and of retention of control would help shed light on the meaning of shift for the two groups. At any rate, in the absence of validating evidence, any conclusions reached here must be tentative. Work is in progress to validate the findings of the present study on a larger group of psychotics selected from patients at a neuropsychiatric hospital.

Summary

In this study a new dimension of the Rorschach, the determinant shift from free association to inquiry, was applied to a group of psychotics and psychoneurotics.

The specific predictions were:

1. Neurotics would show greater shift from free association to inquiry than psychotics.

2. Neurotics would show greater shift on variables presumably indicative of anxiety, *FY* and *Total Y*. Using both the *t* test and the sign test, both predictions were confirmed at very significant levels of confidence.

It was further demonstrated that the Determinant Shift score emphasized differences between the two groups that the classical method of scoring the determinants tended to obscure.

Sixteen psychotic patients who shifted least were compared on the free association scores with 15 psychotic patients who shifted most. The group which showed most shift gave significantly more *M* and *CF*, while the group which showed least shift gave significantly more pure *C*. The findings for psychotics were compared with those for psychoneurotics, and the implications of these findings were discussed.

In view of the significant findings of this and other studies, the Determinant Shift score appears to be suitable for further experimental investigation and validation.

Received August 31, 1953.

References

1. Gibby, R. G. Examiner influence on the Rorschach protocol. *J. consult. Psychol.*, 1952, **16**, 449-455.
2. Gibby, R. G., & Stotsky, B. A. The relation of Rorschach free association to inquiry. *J. consult Psychol.*, 1953, **17**, 359-364.
3. Salk, L. The relationship of elaboration on the Rorschach inquiry to continuance in psychotherapy. Unpublished doctor's dissertation, Univer. of Michigan, 1953.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Rorschach Card Rejection in Psychodiagnosis

Ivan N. Mensh and Joseph D. Matarazzo
Washington University School of Medicine

The phenomenon of rejection of the stimulus cards of the Rorschach test has been observed since Rorschach's early work with this diagnostic method. He noted the differential significance of card rejection by normal, neurotic, "feebleminded hysterical," and schizophrenic subjects:

> *Failure to answer.* Normal subjects almost never fail to give a response for the several plates. Occasionally neurotic subjects fail to answer; this is caused by inhibitions due to complexes. Feebleminded hysterical subjects often reject the figures because they are afraid they will give stupid answers; we deal with an "intelligence complex." Failure frequently occurs in a typical manner in schizophrenics, even in latent or practically recovered cases; they suddenly refuse to answer, even though they have been giving many and good responses all along. The refusal may come with figures not ordinarily considered difficult. The blocking may be insurmountable although in all other kinds of cases an answer may be obtained on coaxing [8, p. 22].

Beck, in his early monograph [1], also summarized the psychological meaning of card rejections by the schizophrenic, depressed, neurotic, and feebleminded, relating the behavior to "blocking, negativism, or stuporous condition" in the first group; to "generalized inhibition and self-depreciation" in the second; to "self-distrust and lack of confidence" in neurotics; and to "criticism" in the feebleminded. Klopfer and Kelley [6] present greater detail and occasional illustrations of card rejection; Bell [3] examines the "main indications of rejections of the various plates"; Yacorzynski [9], Zubin [10], and Zubin and Young [11], among others, make passing reference to the phenomenon of Rorschach card rejection by subjects.

Review of these reference texts in the Rorschach method does not indicate, however, any statistical basis for differential significance of card rejection by the various diagnostic groups seen in clinical practice. Further, a study of the indexes of *Psychological Abstracts* since 1946 and of the *Rorschach Research Exchange and Journal of Projective Techniques* since its first volume yielded but a single article [5] devoted solely to card rejection in Rorschach testing, and this was in a Swiss journal. Although a number of studies have reported card rejection as part of the findings, none has been designed specifically to investigate systematically the frequency of rejection of the stimulus cards. The present report is presented as an attempt to offer data on Rorschach card rejection and its relationship to other personality characteristics.

Research Design

Because of the significance and use of card rejection in interpretation of data gathered by the Rorschach method, the present study was undertaken to investigate the statistical incidence of this phenomenon and to seek for correlative data. The sample of subjects studied was a total sample during a two-year period, and consisted of 201 patients given psychological examinations in the neuropsychiatric service of an inpatient hospital. Of the total sample, 100 were psychoneurotic (*PN*), 74 were psychotic (*PS*), and 27 were brain-damaged (*OR*), as diagnosed, independently of the Rorschach data, by psychiatric and psychological staff members. Data were collected on the incidence of rejection of stimulus cards of the Rorschach, and on other characteristics of the sample studied, including age, sex, Wechsler IQ, education, *R*, *T/1R*, and *T/R*. An analysis of the reaction time characteristics and their significance has appeared elsewhere [7].

Results

Identifying data and Rorschach response characteristics of the subjects studied are summarized in Table 1. The differences

Table 1

Rorschach and Other Identifying Variables among the Three Samples Studied

Variable	Psychoneurotic (*PN*)		Psychotic (*PS*)		Organic (*OR*)	
	Respond	Reject†	Respond	Reject	Respond	Reject
N	76	24	60	14	17	10
Age						
Mean	30.62	30.64	27.28	28.78	37.53	40.89
SD	9.60	9.24	11.25	8.78	15.27	11.14
Sex						
Male	32	12	26	7	10	8
Female	44	12	34	7	7	2
Education						
Mean	13.40**	10.33	12.92	11.08	11.88	8.57
SD	2.79	3.78	3.18	4.01	4.30	1.18
IQ (W-B)						
Mean	113.47**	99.90	104.18	96.00	94.33	90.50
SD	12.07*	19.63	17.57	25.17	20.35*	8.17
R						
Mean	39.20**	16.00	55.47**	19.78	29.88*	13.00
SD	21.14**	7.96	50.94*	7.60	27.09**	4.37
T/R						
Mean	41.54**	65.33	42.17	72.15	72.76	61.00
SD	27.86	31.92	24.42**	65.12	49.64	26.42
T/1R						
Mean	18.00	25.54	16.68*	31.36	28.24	22.89
SD	14.62	19.75	15.10*	25.11	22.78	18.61

* Responding and rejecting samples significantly different at the .05 level.
** Responding and rejecting samples significantly different at the .01 level.
† Patients rejecting one or more cards.

among the three samples with respect to Rorschach response characteristics and the other identifying data were reported earlier [7]. In brief, the mean age of the organic sample was significantly greater and their mean Wechsler-Bellevue IQ was significantly less than the means of the two other groups. These differences may be the result, at least in part, of lowered functioning due to the cerebral pathology of this group. The neurotic and psychotic groups also had more years of schooling. Further, among the patients diagnosed as suffering from organic pathology there were significantly more men than women (67 per cent males, contrasted with 44 to 45 per cent in the other two samples). Thus, the three samples were not matched groups in *N*, age, sex, education, or intelligence level; this raises the problem of significance of variation among the Rorschach characteristics of the samples. Statistically significant differences among the mean values of Rorschach responses may possibly be attributed to one or more of the identifying variables listed above, and not to a disease category. Therefore, a series of correlations was necessary to establish the extent of relationship among the identifying variables and the response variables. These were computed and are reported in Table 2.

It may be seen that in none of the samples did age correlate significantly with other factors, indicating that this variable probably was not significant in the varying Rorschach response characteristics, though the *OR* group was significantly older than groups *PN* and *PS*. However, both education and intelligence level did correlate significantly with certain of the other response characteristics, but there was not a consistent relationship from sample to sample except between education and IQ, and here the correlations were of the same order: .63 to .66. There was a positive correlation between education and total number of responses (*R*) for the *PN* and *OR* groups but not for the *PS* group; yet only in the latter group was there a significant though inverse relationship ($r = -.35$) between education and mean time per first response (*T/1R*).

Table 2

Correlations between Variables among the Three Samples

Variable	Sample		
	PN ($N = 107$)	*OR* ($N = 29$)	*PS* ($N = 74$)
Age – Educ.	—.07	—.16	—.03
Age – IQ	.22	.21	.08
Age – *R*	.00	—.12	—.07
Age – *T/R*	—.03	.30	.20
Age – *T/1R*	.06	.02	.21
Age – Σ*T*	—.02	.25	.06
Educ. – IQ	.63**	.63**	.66**
Educ. – *R*	.28**	.49*	.14
Educ. – *T/R*	—.19	—.15	.00
Educ. – *T/1R*	—.16	—.18	—.35**
Educ. – Σ*T*	.09	.37	.09
IQ – *R*	.32**	.18	.25
IQ – *T/R*	—.29*	—.07	—.11
IQ – *T/1R*	—.23	—.28	—.19
IQ – Σ*T*	.07	.19	.02
R – *T/R*	—.41**	—.30	—.22
R – *T/1R*	—.32**	—.11	—.17
R – Σ*T*	.54**	.74**	.75**
T/R – *T/1R*	.60**	.60**	.70**
T/R – Σ*T*	.34**	.34	.25*
T/1R – Σ*T*	.30**	.35	.13

* Significant at 5% level of confidence.
** Significant at 1% level of confidence.

Similarly, there was no consistency in correlations between IQ and factors other than education, and there were significant correlations only for IQ – *R* and IQ – *T/R* in the *PN* sample. In summary, of the 63 correlations 19 were significant, 16 more at the $p = .05$ level of significance than would be expected from chance variation alone (see Table 2). However, the relationships did not appear consistently, although the relative significance of *R* is apparent in the 7 of the 19 significant correlations in which *R* was one of the variables, and in the 8 correlations in which time was a factor. Further, 10 of the 19 significant correlations occurred in the *PN* group, with the remaining 9 about evenly distributed between the other two groups, suggesting the greater relative order of relationship among these variables in neurotics as compared with the more seriously disorganized individuals of the psychotic and organic pathology samples.

In Tables 3 and 4 are presented the data on Rorschach card rejections in the total sample, tabulated by disease category, sex, card, and individual. Although twice as many organics, proportionally, as psychotics rejected cards (37 per cent and 19 per cent, respectively), and as many as 24 per cent of the sample diagnosed as psychoneurotic, the value of p for the chi-square statistic was between .20 and .10, not significant at the usually accepted level of .05 or lower. The mean number of cards rejected was variously 1.64 (*PS*), 2.38 (*PN*), and 1.90 (*OR*). Among the 201 patients in the total sample there were 99 rejec-

Table 3

Rorschach Rejections by Sample, Sex, and Card

Sample and sex	N	Rejections by cards										Total
		I	II	III	IV	V	VI	VII	VIII	IX	X	
Psychoneurotic												
Male	44	1	3	0	4	2	5	4	1	5	3	28
Female	56	1	2	0	3	2	4	3	1	8	5	29
Total	100	2	5	0	7	4	9	7	2	13	8	57
Psychotic												
Male	33	0	1	0	2	0	3	1	1	4	2	14
Female	41	0	0	0	2	0	2	3	0	0	2	9
Total	74	0	1	0	4	0	5	4	1	4	4	23
Organic												
Male	18	0	0	1	1	1	2	3	0	4	2	14
Female	9	0	0	0	2	0	0	1	0	1	1	5
Total	27	0	0	1	3	1	2	4	0	5	3	19
All groups												
Male	95	1	4	1	7	3	10	8	2	13	7	56
Female	106	1	2	0	7	2	6	7	1	9	8	43
Total	201	2	6	1	14	5	16	15	3	22	15	99

Table 4

Rorschach Rejections by Sample, Sex, and Individual

Sample and sex	*N*	Frequency of rejection						Total persons
		1	2	3	4	5	6	
Psychoneurotic								
Male	44	2	5	1	2	1	0	11
Female	56	6	2	3	1	0	1	13
Total	100	8	7	4	3	1	1	24
Psychotic								
Male	33	4	0	2	1			7
Female	41	5	2	0	0			7
Total	74	9	2	2	1			14
Organic								
Male	18	4	0	2	1			7
Female	9	2	0	1	0			3
Total	27	6	0	3	1			10
All groups								
Male	95	10	5	5	4	1	0	25
Female	106	13	4	4	1	0	1	23
Total	201	23	9	9	5	1	1	48

tions of stimulus cards by 48 individuals, a mean rate of 2.06 for those who did reject. These rejections were principally of cards IV, VI, VII, IX, and X. Similarly, Klopfer and Kelley [6] reported rejections most often of cards II, IV, VI, and IX. In the present study 24 per cent of the total sample rejected cards, 50 per cent of the stimulus cards provoked 83 per cent of the rejections, and in these rejections 23 of the 48 patients rejected only one card and only two rejected more than four cards. Brockway *et al.* [4] reported rejection by 22.5 per cent of their neurotic patient sample, an incidence consistent with our findings. Hiltmann [5] also concluded from a review of Rorschach literature that card rejections "are rather frequent."

Tests of significance of differences between means and their sigmas for responding and rejecting patients are summarized in Table 1, which also reported the sample characteristics. The "reject" columns represent those patients who rejected one or more stimulus cards. There were no significant differences between the subsamples, within each diagnostic category, with respect to age and sex. On the education variable, only the *PN* group showed higher education for the responding sample than for those rejecting Rorschach stimulus cards. Since education and Wechsler-Bellevue IQ were correlated significantly (r's of .63 to .66) for each of the three diagnostic samples, only among the *PN* group did the responding patients have higher IQ's, as would be expected. However, variability differed significantly among both the *PN* and *OR* samples, but in reverse directions — greater variation among those *PN* patients rejecting cards, but less variation among rejecting *OR* patients. As would be expected from the consequences of card rejection, for all samples there were significantly fewer responses and correspondingly decreased variation among the patients rejecting cards. Yet, only in the *PN* sample was there a significant mean difference for T/R. The patients of the *PS* sample who rejected cards were far more heterogeneous with respect to the T/R factor, so that the apparent mean difference of 30 seconds was not significant. But in this sample alone were there significant differences between means and sigmas in the analysis of data on the $T/1R$ variable.

Summary

Rorschach card rejection has been observed and reported since the introduction of the Rorschach technique for psychodiagnosis. Differential significance of the phenomenon of card rejection frequently has been suggested as a diagnostic aid, but a systematic study of the statistical incidence of this behavior has not been reported. In the present study, 201 hospitalized patients were examined with the Rorschach method. In the total sample there were 100 psychoneurotic, 74 psychotic, and 27 brain-damaged patients. The three groups were not matched but represent a typical sample of patients seen in clinical psychologic practice on an inpatient service. Intercorrelations among variables representing sample characteristics were not consistent, but suggest that the relative order of relationships among these characteristics in neurotics is higher than in the samples of the more seriously disorganized patients diagnosed as psychotic or brain-damaged.

Rejection of one or more of the Rorschach stimulus cards was found in 19 to 37 per cent of the samples; five of the ten cards provoked 83 per cent of the rejections; in the responses of subjects who rejected the cards, on the average, two cards were rejected per patient, although about half of the rejecting individuals rejected but one card and in only two cases were more than four cards rejected. A number of trends that differentiated patients responding to all of the stimulus cards from those rejecting one or more of the cards appeared, but there were few statistically significant differences. As in an earlier study, the major finding with respect to Rorschach card rejection is the extreme variability in response characteristics, emphasizing again the care necessary in utilizing data gathered by the Rorschach method to interpret behavior.

Received October 20, 1953.

References

1. Beck, S. J. *Introduction to the Rorschach method.* New York: Amer. Orthopsychiat. Ass., 1937.
2. Beck, S. J. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune & Stratton, 1944.
3. Bell, John E. *Projective techniques.* New York: Longmans, Green, 1948.

4. Brockway, Ann L., Gleser, Goldine C., & Ulett, G. A. Rorschach concepts of normality. *J. consult. Psychol.*, 1954, **18**, 259-265.
5. Hiltmann, Hildegard. Das Versagen im Rorschachtest. (Rejections in Rorschach testing.) *Beih. Schweiz. Z. Psychol. Anwend.*, 1948, No. 13, 123-130. (*Psychol. Abstr.*, 1948, **22**, No. 3941.)
6. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1946.
7. Matarazzo, J. D., & Mensh, I. N. Reaction time characteristics of the Rorschach test. *J. consult. Psychol.*, 1952, **16**, 132–139.
8. Rorschach, H. *Psychodiagnostics.* (3rd Ed.) New York: Grune & Stratton, 1942.
9. Yacorzynski, G. K. *Medical psychology.* New York: Ronald, 1951.
10. Zubin, J. *Quantitative techniques and methods in abnormal psychology.* New York: Columbia Univer. Bookstore, 1950.
11. Zubin, J., & Young, K. M. *Manual of projective and cognate techniques.* Madison: College Typing Co., 1948.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Intelligence Test Reliability for Cerebral Palsied Children[1]

David H. Crowell
University of Hawaii
and Doris C. Crowell
Honolulu, T. H.

Cerebral palsied children having at least two psychological examinations using the same instrument for the initial and final contact, and having mental ages definitely stated, were selected from the register of the Board of Health, Honolulu, Hawaii, and the Psychological Clinic, University of Hawaii. Individual diagnoses of the 61 cases selected included the full range of motor disabilities as well as muscular involvements. Examinations, ranging from 2 to 6 per individual, were based upon the Stanford-Binet or Porteus Composite Binet [1], 49 cases; the Cattell Infant Intelligence Scale, 11 cases; and the Merrill-Palmer Scale of Mental Tests, 1 case. Seventeen different psychologists were involved in 46 cases and the same examiner in 15 cases.

Based upon the median of all test-retest intervals, two groups were formed: Group I, 16 males and 12 females with an average test-retest interval of 15 months, and Group II, 20 males and 13 females with an average test-retest interval of 43 months. The total group had an average test-retest interval of 30 months.

This sample was highly skewed, with roughly 50 per cent of the group classifiable as severely retarded. Analysis of the total retest results showed that approximately 51 per cent of the total group deviated within ± 5 IQ points of the initial test score and approximately 75 per cent within ± 10 IQ points of the initial test score. Pearsonian coefficients of reliability, based upon mental ages, and confidence limits at the 1 per cent level are as follows: Group I, $r = +.98$ with confidence limits of +.94 and +.99; Group II, $r = +.78$ with confidence limits of +.52 and +.91; and total group, $r = +.92$ with confidence limits of +.87 and +.95. Standard errors of estimate are, respectively, ± 7, ± 13, and ± 11.

The data suggest that the mental age obtained on the initial examination of a cerebral palsied child can be accepted as highly reliable for ordinary diagnostic purposes, provided the clinical psychologist has adequate training and experience. If errors of clinical judgment on the initial examination occurred, they erred on the side of caution. Moreover, these coefficients of stability are highly significant and have wide generality since there were increased chances for the operation of errors inherent in the measuring instrument, errors due to clinical judgment, and errors resulting from fluctuations in ability because of maturation.

Brief Report.
Received March 1, 1954.

[1]An extended report of this study may be obtained without charge from David H. Crowell, Department of Psychology, University of Hawaii, Honolulu 14, Hawaii, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4287 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for 35 mm. microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Reference

1. Jorgensen, Margaret. A comparative study of the Stanford-Binet Form L and the Porteus Composite Binet. Unpublished master's thesis, Univer. of Hawaii, 1951.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

A Cross Validation of Davidson's Rorschach Adjustment Scale

Raymond J. Corsini
University of Chicago
and Harold F. Uehling
Wisconsin Department of Public Welfare

Among the objective scales based on the Rorschach protocol are those constructed by Buhler [1], Davidson [2, 3], Munroe [9], Hughes [6], and Piotrowski [10]. The first three are of particular interest since they attempt to measure general adjustment. These scales were constructed along the general principles employed by objective test constructors: two populations believed to be different in terms of a criterion were established; various signs were tested to determine their discriminative powers; signs that met certain limits of significance of difference were isolated, and were then combined to form the scale. However, the crucial test of the value of any scale is its ability to discriminate between new populations.

The purpose of this study is to test the Davidson scale for generality of measurement. The Davidson scale was selected for several reasons: first, the signs are limited to 17, 15 of which are so objective that a clerk might score them; second, previous research has been done with this scale to indicate that it is reliable [2, 3]; third, the scale has been demonstrated capable of differentiating between "adjusted" and "maladjusted" school children [2, 5, 8].

The objectification of the Rorschach by lists of signs has two purposes: (*a*) validation of hypotheses relating to various signs, and (*b*) the possible scaling of various degrees of emotional maladjustment. Most Rorschach workers are generally dubious about the value of such scales, believing that projective instruments reveal unique and ineffable aspects of personality and that the quantitative conversion of qualitative aspects of the personality yield little of extra value. However, it is possible that objective procedures have some value. For example, in comparing two clear-cut populations, one "normal" and one "abnormal," if it is found that differences with regard to any sign are nonsignificant or even in a direction not predicted, then the clinical insights that suggested the interpretation of these signs may be questioned.

Method

The two populations are called "Guard" and "Inmate." Both consist of males, matched for age. Since it was not possible to match for intelligence, only protocols of inmates who were known to have at least average intelligence were selected; also the guards had the equivalent of high school graduate standing, and were presumably at least average in intelligence. The guard population is composed of men accepted as probationary prison guards at the Wisconsin State Prison at Waupun. They ranged in age from 21 to 45, and were of undamaged reputation. Each probationary guard had passed, in addition to the usual civil service requirements, the criterion of an interview by the appointing authorities. Therefore, they form a screened population since people who would appear unsuitable in terms of personality defects or handicaps of adjustment would not be hired. No evidence exists to support a belief that this group as a whole differs from any normal sample of the adult male population.

The inmate population consists of men committed to prison for felony crimes. This alone might be considered to be presumptive evidence that the group as a whole is abnormal, but the population actually subjected to the Rorschach was a selected sample of this larger group. Only those inmates who, in the opinion of a

psychiatrist, appeared seriously emotionally conflicted were referred to the psychologist for projective analysis. Therefore, the sample of inmates who were matched for age against the guards was doubly distilled for abnormality: first by the court, and second by the initial psychological and psychiatric examinations.

All Rorschachs in this study were administered by Uehling. Analysis was done by both writers, and for the two signs that called for judgment, *no color shock* and *no shading shock* (signs 15 and 16), independent evaluations were made. If agreement did not occur, the protocols were examined more carefully until an agreement was reached. The analysis of the protocols was done without knowledge of the class identity of the subject. The frequency of occurrence of the Davidson signs was tabulated for each population. The significance of the difference was obtained by dividing the percentage difference by the standard error of the percentage of the difference from Edwards' formula 21 [4, p. 77]. The resulting *t*'s were converted into fiducial limits of probability using the two-tailed test on the assumption that none of these signs is significant in either direction [7]. The mean and standard deviation of the scale for each of the populations were also calculated.

Table 1
Occurrence of Favorable Signs for Two Populations of Prison Inmates and Guards ($N = 50$ for each population)

Sign	Inmates	Guards	*t*	*p*
1. $M > FM$ or $M = FM$	26	27	.02	.9
2. *M*, 3 or more	8	13	1.23	.3
3. *Sum* $C > Fc + c + C'$	30	21	1.80	.1
4. *F%* — 50 to 20	21	18	.75	.5
5. $Dd + S$ 10% or less	32	34	.04	.9
6. *P*, 4 or more, up to 30% of *R*	33	38	1.11	.3
7. *R* more than 20	25	30	1.01	.4
8. $FC > CF$ or $FC = CF$	25	26	.02	.9
9. *FC* 2 or more	21	30	1.80	.1
10. No pure *C*	23	31	1.61	.2
11. *%8*, 9, 10-40 to 60	17	14	.65	.6
12. $FK + Fc$, 2 or more	12	11	.02	.9
13. $W:M = 2:1$	12	9	.75	.5
14. *A%* 50 or less	30	37	1.49	.2
15. No color shock	23	31	1.61	.2
16. No shading shock	37	34	.75	.4
17. No refusals	35	45	2.50	.02

Results

Table 1 lists the 17 signs, the frequency of their occurrence in both groups, the *t* tests, and the levels of significance. It will be noted that 14 of the 17 signs are in the expected direction. However, the "abnormal" population does "better" than the "normal" population on signs 3, 4, and 16. Sign 3, *Sum C greater than Fc plus c plus C'*, is significant at the 10% level. Of the remaining 14 signs, twelve are insignificantly in favor of the "normal" group. Sign 9, *FC two or more,* is significant at the 10% level, and sign 17, *no refusals,* is significant at the 2% level.

The mean Davidson score for the 50 inmates is 8.24 ± 2.32, and for the guards it is 9.08 ± 2.36. The difference of 0.84 is significant at the 10% level with a *t* of 1.77. These figures compare favorably with the scores reported by Davidson for various populations [3]. The only adult norm is for seven college students, given as 8.14; this score is nonsignificantly lower than both our inmate and guard norms.

Discussion

Scales may be restricted to special populations. While a scale for length is general for any object, psychometric scales may be specific. Thus an intelligence test may be suitable for only a particular population, and scales for measuring general adjustment with the Rorschach may be restricted in their applicability. In the case of the Davidson scale as applied to our populations, 14 of the 17 signs are in the expected direction, but of the three that meet the 10% level of significance, one of them is in the wrong direction. Only one sign, number 17, *no refusals,* was significant at beyond the 5% level. Whereas only 10% of the guards had any refusals, 30% of the inmates refused one or more cards.

The scale as a whole measures in the expected direction, but not to any significant degree. The gain resulting from using this scale is minimal. We could expect to select eight fewer guards and eight more inmates out of a population of 100 guards and 100 inmates than by chance if we were to use this scale for selection purposes. This is certainly no impressive saving over chance.

Future Research

In a previous study [12] using the same populations, it was discovered that while the reaction times of the inmates and the guards did not differ significantly from each other, both differed significantly from the reaction times reported by Sanderson [11]. The reason for the difference was stated to be the reluctance of the guards and inmates in contrast to the eagerness of Sanderson's group. However, this probable reluctance should not be germane to the present investigation.

It is intended in the future to check the discriminative capacity of other scoring schemes to separate these populations; and, in addition, to determine the ability of psychologists, unaided, to separate the two groups.

Summary

The Rorschach protocols of two populations equated for sex, color, age, and intelligence were analyzed in terms of the signs on the Davidson Rorschach Adjustment Scale. One population consisted of probationary prison guards who took the Rorschach as part of the routine of selection and are considered to be "normal." The other population, consisting of prison inmates convicted of serious crimes, screened by a psychiatrist, and referred to projective testing in terms of suspected personality difficulties, is considered "abnormal."

Of the 17 Davidson signs, only 3 were significant beyond the 10% level, 1 of them in the "wrong" direction. Only one sign, *no refusals,* was significant beyond the 5% level. The scale as a whole was significant at the 10% level. By setting a cutting score at 9.00, an 8% saving over chance resulted, which is believed to be too small to warrant the use of this method for the purpose of separating adults in terms of personality maladjustment.

Received November 30, 1953.

References

1. Buhler, Charlotte, Buhler, K., & Lefever, D. W. *Rorschach standardization studies. I. Development of the basic Rorschach score with manual of directions.* Los Angeles: Rorschach Standardization Studies, 1948.
2. Davidson, Helen H. *Personality and economic background.* New York: King's Crown Press, 1943.
3. Davidson, Helen H. A measure of adjustment obtained from the Rorschach protocol. *J. proj. Tech.,* 1950, **14,** 31-38.
4. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
5. Gair, M. Rorschach characteristics of a group of very superior seven-year-old children. *Rorschach Res. Exch.,* 1944, **8,** 31-37.
6. Hughes, R. N. Rorschach signs for the diagnosis of organic pathology. *Rorschach Res. Exch.,* 1948, **12,** 165-167.
7. Jones, L. V. Tests of hypotheses: one-sided versus two-sided alternatives. *Psychol. Bull.,* 1951, **49,** 43-46.
8. Margulies, H. Rorschach responses of successful and unsuccessful students. *Arch. Psychol., N. Y.,* 1942, No. 271.
9. Munroe, Ruth L. The inspection technique: a method of rapid evaluation of the Rorschach protocol. *Rorschach Res. Exch.,* 1944, **8,** 46-70.
10. Piotrowski, Z. A. The Rorschach ink-blot method in organic disturbances of the central nervous system. *J. nerv. ment. Dis.,* 1937, **86,** 525-537.
11. Sanderson, H. Norms for "shock" in the Rorschach. *J. consult. Psychol.,* 1951, **15,** 127-129.
12. Uehling, H. F. Rorschach "shock" for two special populations. *J. consult. Psychol.,* 1953, **16,** 224-225.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

Failure to Return for Treatment in Tested and Nontested Clinic Patients[1]

Miriam McCue, Morris Goodman,[2] and Melvin Rosenthal[3]
VA Mental Hygiene Clinic, Lowell, Massachusetts

When new clients fail to return to a clinic for treatment after they have been psychologically tested, it is not uncommon to assume that the tests were "too threatening."

To test the validity of this assumption, the following hypothesis was formulated: Giving psychological tests to patients when they first come to a clinic has the effect of increasing the incidence of failure to return for treatment.

A brief battery of tests was given to all clients who reported to an out-patient clinic during a six-month period. The tests were the Kent EGY, Draw-A-Person, Bender Gestalt, and a sentence completion test. The *Ss* who had received therapy since discharge from the service were eliminated, as were Korean and World War I veterans in the final tabulation. The experimental group contained 20 *Ss*.

Patients who were not tested at the time of intake made up the control groups. There were 51 *Ss* who entered during the same months of the previous year, and 10 who entered during the same months of the year following that selected for "intake testing."

The *Ss* in all groups were male, white World War II veterans who lived in the city in which the clinic is located, or in nearby communities. Their ages ranged from 23-43.

Only one (5%) of the tested *Ss* did not return for treatment, whereas in the two control groups six and one *Ss*, respectively, i.e., 11% of the combined control group, did not return. Thus the hypothesis is not borne out. Fewer of the tested *Ss* than of the nontested *Ss* failed to return for treatment. The difference is not significant, however, and the direction it takes could have occurred by chance variation in the sampling of patients.

Length of time in treatment may be influenced by many factors, of which testing is only one. If the testing situation was threatening to patients when they first came to the clinic, the *Ss* in the experimental group should have shown a tendency to remain in therapy a shorter time than the control *Ss*. An incidental finding of this study does not support this second hypothesis. The median number of times the tested *Ss* returned for treatment was 10, while the medians were 8 and 7.5 respectively for the other two groups. Again the trend seems to be the reverse of the hypothesis. However, the difference between the control and the experimental group medians was not significant, and therefore time in treatment, in terms of therapeutic interviews, does not differ significantly for the tested and the nontested *Ss*.

Apparently the testing at this clinic was not threatening enough to deter the patients from returning for treatment, or to reduce the number of their visits. The effects of different tests, different examiners, and different therapists remain problems for further research.

Brief Report
Received May 3, 1954

[1] An extended report of this study may be obtained without charge from Miriam McCue, 20 North Hancock Street, Lexington, Mass., or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4251 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[2] Now at VA Mental Hygiene Clinic, Newark, New Jersey.

[3] Now at VA Hospital, Brockton, Massachusetts.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

An Evaluation of Psychological Reports In a Clinical Setting[1]

Sol L. Garfield
VA Hospital, Downey, Illinois
Ralph W. Heine and Morton Leventhal
VA Mental Hygiene Clinic, Chicago, Illinois

The clinical psychologist currently spends a large share of his time administering psychological tests to patients and writing fairly extensive reports of his findings. These reports are usually presented at some form of staff conference and later are incorporated into the case file of the patient, where they become a permanent part of his record. Although the psychological test report represents the final summary of the psychologist's clinical examination and is viewed as his unique contribution to the total diagnostic study of the patient, comparatively little discussion has appeared in the literature concerning this important activity. Furthermore, while a few accounts concerning the nature and form of psychological reports have appeared [1, 2, 3], critical evaluations of current report writing have been negligible.[2] Because of this lack in a significant area of the psychologist's professional work, it was deemed worth while to report on the attempts of one clinic to evaluate critically the psychological test reports used in the daily activities of the clinic.[3]

The Present Study

The need for the present study was recognized as a result of the regular interdisciplinary staff discussions at the clinic. In the course of these meetings, various dissatisfactions with the psychological test report were voiced. The most important criticisms dealt with the meaningfulness of the psychological report and its general usefulness. As a result of these staff discussions, the psychologists were stimulated to explore the problem at greater length, and a more formal investigation was undertaken in the hope that a better evaluation of the problem and some specific guides toward improving the psychological reports might be obtained.

Procedure

The procedure eventually decided upon was to select a one-month period far enough in the past to avoid ready identification of the author, but recent enough to be representative of current psychological reports. All the reports written during this month, 16 in number, were stripped of all identifying data and constitute the material upon which the present study is based.

The reports were then numbered from 1 to 16. The numbers were placed in a hat and one of them was selected. Copies of the report bearing this number were then typed and distributed to every full-time professional staff person in the clinic. In addition, the remaining 15 reports were retyped and distributed in a random fashion to the clinic staff. In this manner, each staff person was given two psychological reports in a blank envelope and asked to evaluate each of the reports and assign a letter grade to it. The evaluations were to be done anonymously and returned to a secretary. The only identification on the envelope related to the professional group to which the staff member belonged. In this way we were able to compare the evaluations of psychiatrists, psy-

[1] From the VA Hospital, Downey, Illinois and the VA Regional Office, Chicago, Illinois.

[2] The first book on writing clinical reports [1] was published as this manuscript was being prepared.

[3] The project reported here was a cooperative one in which all staff members of the VA Mental Hygiene Clinic, Chicago, Illinois, participated. The authors wish to acknowledge this participation and to thank especially Drs. D. G. Wright, John G. Watkins, and Miss Clara Weimer for their aid in completing the project.

chologists, and social workers as a group. Specific written instructions were included as follows:

> Your cooperation and participation is requested in this survey (anonymously). The Psychology Department believes that by securing independent evaluations from the various members of the staff regarding psychological reports, it will be possible to incorporate many valuable suggestions to the end of better service in the diagnostic area.
>
> In order that there may be specific remarks with which we can work, each participant in this survey is being supplied with two sample reports, one of which is being evaluated by everyone. There are wide margins where any notations or remarks may be made regarding the adjacent paragraphs, phrases, etc. The remarks may cover any aspect, from theory to grammar. In fact, feel free to cover any range of comment you wish.
>
> By way of explanation, the reports used have been taken from consecutive referrals over a one-month period and have been written by all members of the department who were examining during that period. The numbers on the reports are merely for purposes of tracing them back to the original testing material.
>
> Please give a letter rating at the top of the first page to indicate the over-all quality of the report, e.g., A, excellent; B, good; C, fair; D, poor; F, very poor.

The reports were then given to 9 psychologists, 13 psychiatrists, and 11 psychiatric social workers. All of the staff with the exception of one psychiatrist returned the reports with their written evaluations. Thus, there were 32 sets of evaluations. No attempt was made to ask specific questions or to have the participants rate the reports on selected variables. Rather, we were interested in the spontaneous and free expression of opinion on any aspect of the psychological test reports that might stimulate response, and wanted to leave the matter as unstructured as possible. Although this procedure makes the analysis of the data more difficult, such an approach was deemed advisable in terms of the preliminary nature of the investigation.

Once the evaluations were collected, the writers were selected as a research team to analyze the data. Each of the three authors studied the common report independently; he recorded the comments and tried to order the material in some meaningful manner. Next, several meetings were held to discuss the preliminary results obtained and to arrive at a uniform method of recording the responses to the test report. A check list consisting of six categories with 24 items was then prepared to use in reordering the data in a uniform and meaningful framework. The common report was again reappraised independently by all three investigators in terms of this guide. It was decided at that time also to record a specific comment or criticism only once for each person. For example, if a participant made the same criticism of one report a number of times, it was recorded only once. This step was judged necessary to compensate for the variability among the written comments of individual staff members and as a check against distorting the significance of specific comments. Thus, if one person made the same criticism 10 times, it would receive the same weight as a general criticism made once by each of 10 different individuals. As a consequence, the unit used was the number of individuals making the comment rather than the actual number of times a criticism was made.

After the second independent appraisal of the responses to the common report, the writers reassembled to discuss their findings and plan for the final recording of the data. The final recording and tabulation of responses to all reports was then made by the writers as a team.

Results

The letter grades given to the common report are summarized in Table 1. In general,

Table 1

Grades Given to Psychological Test Reports

	Common report		
Grade	Psychol.	Psychiat.	Soc. wkrs.
A	1	1	5
B	3	7	3
C	5	1	3
D & below		2	
Total	9	11*	11
	Other reports		
Grade	Psychol.	Psychiat.	Soc. wkrs.
A	2	0	0
B	2	1	8
C	4	7	2
D & below	1	4	1
Total	9	12*	11

* One psychiatrist submitted written comment but omitted a grade for the common report.

the report was judged to be satisfactory or better. Social workers were inclined to rate the report much higher than representatives of the other two clinic disciplines. While the median rating of the psychiatrists was higher than that of the psychologists, the only participants who gave the report a rating below C were psychiatrists. This finding may indicate a greater variability in the evaluative attitudes toward psychological test reports held by psychiatrists.

The grades given for the other reports differ in some ways from the distribution obtained for the common report. In general, the social workers and psychiatrists gave lower grades to these reports than to the common report, whereas the psychologists' grades were essentially similar in both instances. Since there were some differences in these reports, no firm conclusions can be reached, but certain trends are discernible. In terms of all the reports evaluated, the psychologists demonstrated the greatest stability of ratings, the social workers tended to give the most positive evaluations, and the psychiatrists the most critical evaluations. This trend was also apparent in the qualitative comments made for each of the test reports.

As indicated previously, some difficulties were encountered in trying to order the spontaneous comments of the staff with regard to the psychological test reports. We purposely wanted a free expression of opinion. However, since the verbatim responses varied greatly, some way to group the data was necessary. The particular classifications finally used are given in Table 2. Attempts were made to use the actual verbalizations and meanings as subheadings as much as possible to avoid misinterpretations of the data. Undoubtedly some of the categories overlap, and some of the different groupings may in part reflect specific orientations of each professional discipline.

As we have organized the data, the critical comments fall into four major categories: (*a*) lack of behavioral referents or supporting data for statements contained in the reports; (*b*)

Table 2

Evaluative Responses to Psychological Test Reports

Categories of evaluative responses	Common test report				Other test reports			
	Psychol. ($N = 9$)	Psychiat. ($N = 12$)	Soc. wkrs. ($N = 11$)	Total ($N = 32$)	Psychol. ($N = 9$)	Psychiat. ($N = 12$)	Soc. wkrs. ($N = 11$)	Total ($N = 32$)
Lack of supporting data or behavioral referents								
No behavioral referents	7	2	1	10	5	1	3	9
Too speculative	0	6	3	9	2	9	3	14
No supporting test data	0	3	0	3	1	8	3	12
Poor expression								
Use of clichés, jargon, etc.	4	1	1	6	5	7	3	15
Loose use of terms	4	2	0	6	5	5	0	10
Vague and unclear	4	0	0	4	4	9	3	16
Poorly expressed	4	0	0	4	2	2	1	5
Poor organization								
No integrated picture of patient	3	0	0	3	2	1	1	4
Poorly organized	0	0	2	2	3	2	2	7
Miscellaneous								
Inconsistent	3	4	1	8	4	8	0	12
Incorrect use of theory	3	3	1	7	0	3	0	3
Test data not differentiated from other data	1	1	2	4	0	2	0	2
Makes unconscious sound conscious	2	2	0	4	1	0	0	1
Doesn't answer referral problem	0	0	2	2	2	1	1	4
Too long and irrelevant	0	0	0	0	4	4	4	12

general lack of clarity of expression; (*c*) poor organization; and (*d*) miscellaneous observations. The first two categories appear to be the most significant in terms of the qualitative comments received. All of the categories, however, will be discussed briefly.

The psychologists reacted most critically to a lack of specific behavioral referents, i.e., individualized examples of very general statements made in the test reports. For example, if a patient was characterized as anxious, depressed, or immature, they wanted to know how this was manifested behaviorally or what the significance of the trend was for the specific patient. Two excerpts from one of the test reports that exemplify this deficiency are the following:

> He has little ability to delay gratification through the medium of mature fantasy activity and thus resorted to immature, wishful inner living.
>
> There is little hazard of the patient acting out since there is a paucity of inner resources.

These statements, as they stand, defy easy translation into terms readily referable to behavior, irrespective of the professional background brought to them by the reader. Thus, it is not surprising that such expressions in a psychological report come under criticism. One of the basic questions — "does this man have strong impulses to commit acts of violence and, if so, how does he cope with them?" — remains unanswered. In the opinion of the writers, the report properly should include a description of the kinds of provocation which evoke hostile aggressive impulses and the ways in which angry feelings are handled by the patient short of directing destructive actions against himself or others. The reader would then be in a position to evaluate along with the diagnostician the "hazard" of this patient's "acting out."

While the psychologists were concerned with the problem of behavioral referents, the psychiatrists reacted to many general statements of personality dynamics with obvious skepticism. Over half of them felt that statements in the test reports were "too speculative" or lacked supporting data. Typical comments in this general category were the following: "I object to this immediate plunge into interpretations, speculations and conclusions without citing test data highlights for each of the tests administered on which to base opinions." "On what is this speculation based? . . . Data, at least some, are essential to evaluate this speculation." "What evidence is there for this interpretation?" "There is a tendency too often to accept certain hypothetical dynamics as fact and base conclusions on such theoretical conclusions." "Again, suggest at least some basic data as obtained and separate from interpretaton and/or speculation." One final illustration is offered as an example of what we considered a conscientious appraisal of the report:

> This portion gives the impression of being based on test results; however, the following fact does not leave such an impression but suggests instead that test material was used as a take-off point for a flight into the type of speculation that suits the examiner's system of psychological thought. Perhaps the means to avoid such impression might be the use of examples from the test material wherever possible, e.g., "as evidenced by," "as reflected in," etc., to distinguish the demonstrable from the supposed.

There was little question that the psychiatrists judged these kinds of criticisms to be among the most important. Such criticisms were also offered by approximately one fourth of the social workers. The reasons for their skepticism are not readily discernible, but it is worth "speculating" about them. It may be that the reports appeared to be too inclusive and that they doubted that such complete portrayals of personality could be obtained solely from test data. Another consideration may be the tone of finality evident in many reports which also tends to stir negative reactions. Undoubtedly, other more obvious factors are also involved. Very evident inconsistencies, statements which contradict other data, and similar factors may have tended to undermine confidence in the reports and to cause one to view them with suspicion. Whatever the factors may have been, the net result was to limit the acceptance of the test reports.

The criticism of lack of supporting data was also accepted by many psychologists with some reservations. It is often difficult, if not impossible, to present evidence succinctly since it comes from a variety of test data. Moreover, psychologists may have some legitimate doubts that members of allied professions are equipped

to interpret the evidence which is presented. The solution of this problem may depend on a general improvement in psychological reporting with perhaps some supporting evidence for certain kinds of inferences.

The second major category of critical responses included comments pertaining to poor and careless expression. The following are representative comments: "vague, unclear, loose use of terms, pure jargon, don't understand, poorly expressed, etc." While this category conceivably overlaps with statements in the first major subdivision, in general the emphasis in the second grouping was on lack of clarity in presentation, poor communication, and the use of technical jargon.

There was no lack of examples of vagueness, obscurity, and overdependence upon clichés in the reports examined. The following are quotations in which the "sins" of the psychologists are self-evident:

1. Hostile trends are strong. He employs reaction formation as a means of control but there are strong residual tensions which strive for access to the motor sphere.

2. He has little internal resources remaining, along with heightened dysphoria and anxiety.

3. There is evidence of some dissociative disturbance of genital sexuality.

4. He attempts to be orderly, to draw closure on his experiences which tends to limit spontaneous activity. Hysterical qualities characterize his behavior in that he has regressed psychosexually, is emotionally over-active on a superficial level and has repressed most unacceptable impulses.

In the third category, "poor organization," the majority of critics referred to the following: (*a*) a failure on the part of the diagnostician to discuss related findings in the same paragraph or section of the report; and (*b*) failure to construct the report in such a way that various aspects of the personality discussed separately were finally integrated into a unified picture of the functioning of the patient's personality. Psychologists, primarily, were critical of this latter aspect of the reports.

In the final category of miscellaneous items we included: (*a*) observations of contradictory statements which occurred not only within the same report but occasionally within a single sentence; (*b*) instances where the diagnostician clearly did not understand the personality theory that he was employing as a guide for his interpretations; (*c*) comments noting the inclusion of what was obviously social history data as a test finding; (*d*) statements concerning the length of reports and the failure to answer the referral problem; and (*e*) criticisms about "making the unconscious sound conscious." The latter referred to personality descriptions that sounded as if the individual patient consciously planned his behavior in the manner described, e.g., "since he could find no mature means of gratifying his needs, he resorted to somatization as a means of obtaining gratification."

The matter of inconsistency was not limited to any specific area, but in some cases was a reflection of an uncritical use of test terminology and clinical concepts. An example of the latter is evident in the following comment: "How can he [the patient] be 'affectively impoverished' and at the same time show 'heightened dysphoria' and anxiety — do not the latter indicate capacity for affect?"

The critical comments on the length of reports were made by approximately one third of the staff in reference to reports other than the common one. Whereas the common report was two pages in length, the other reports varied from three to five pages, and this fact may account for the relatively lower ratings given to these reports. These findings would imply that a concise report more specifically directed to the referral problem is much better received than a long rambling one that attempts to present an all-inclusive personality description.

Finally, it should be mentioned that there were, as might be expected, scattered criticisms of grammar, "English," and sentence structure. These, however, were sufficiently limited in number as to warrant bare mention here. Also, a number of favorable comments were made, largely by the social workers. This finding seems consistent with their tendency to be less critical and in general to give higher evaluations to the reports. The positive comments were such as to add weight to the categories of criticism shown in Table 2; that is, reports were viewed favorably for being "concise," "clear," "well written," and directed to answering the referral question. The term

"good" was used relatively often with no specific reason given for its use. Perhaps one quotation might help indicate some of the bases for favorable reaction.

> I feel that this is an excellent type of report principally because the language is *simple*, i.e., nontechnical, and the construction is direct and to the point. The "lay-out" too is good.

Summary and Conclusions

A one month's sample of psychological test reports was distributed to all professional staff members in the mental hygiene clinic for their critical evaluation. All identifying data on the reports were removed, and the evaluations were done anonymously. The authors of the study first analyzed the evaluative comments separately and worked out a list of categories which would facilitate the ordering of the data and yet retain the emphases given by the individual staff members. Later the authors worked together on the final tabulation and appraisal of the data.

Some differences between the three professional disciplines were noted. In general, the psychiatrists tended to be most critical of the reports, while the social workers were least critical. The psychiatrists most frequently criticized the reports for being too speculative, omitting supporting data, being vague, unclear, and inconsistent. The psychologists were primarily critical of the lack of behavioral referents for generalized statements made in the reports. They were also sensitive to the use of clichés, technical jargon, and the lack of clarity in several of the reports. In general, the main criticisms of the reports were related to the lack of supporting data or behavioral referents and to a lack of clarity and concise communication.

On the basis of this pilot study, it appears that if psychological test reports are to be more useful they should be as concise as possible, avoid generalities and the use of technical jargon, give supporting data for certain kinds of inferences, avoid "speculative dynamics," and try to give a portrayal of a specific individual using behavioral terms wherever possible. Partly as a result of this study, the staff psychologists at the Chicago VA Mental Hygiene Clinic are revising their style of reporting psychological test data.

Received November 5, 1953.

References

1. Hammond, K. R., & Allen, J. M., Jr. *Writing clinical reports.* New York: Prentice-Hall, 1953.
2. Holzberg, J. D., Alessi, S. L., & Wexler, M. Psychological case reporting at psychiatric staff conferences. *J. consult. Psychol.*, 1951, **15**, 425–429.
3. Sargent, H. D. Psychological test reporting: an experiment in communication. *Bull. Menninger Clin.*, 1951, **15**, 175-186.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Some Common Misconceptions About Neuroticism[1]

Harrison G. Gough
University of California, Berkeley

A respectable science of psychopathology needs to be based upon, among other things, a thorough and accurate knowledge of what particular beliefs, values, self definitions, attitudes, and self conceptions are characteristic of persons diagnosable as maladjusted. Theories about process, about the "self," about motivations, and about therapeutical techniques can hardly be evolved without some minimum pre-establishment of "facts" in this realm. That is, for example, if one is to develop a theory to account for the contention that most neurotics say that they get a "raw deal" from life, then we must first make sure that most neurotics *do,* in fact, express this belief.

It is rather surprising, if one reviews the psychological literature of the past 20 years, to discover how little has been done to determine the characteristic and differentiating beliefs and attitudes of persons classified with respect to their psychological adjustment. We have some general notions, for example, that neurotics feel misunderstood, that they often have bad dreams, that they often expect to fail in what they do, that they desire more sympathy than they get, that they find it hard to concentrate, and so forth. Yet, in spite of the diffusion of ideas such as these, there are only a few studies which have actually asked patients for responses on these issues and then compared the frequency of their agreements with those of control groups.

The few studies done, such as those of Page, Landis, Katz, and Zubin [6, 7, 10], and Hathaway and McKinley [4, 5, 8], have revealed findings strikingly discrepant from the pervasive stereotypes of both professional and lay persons. The clinical stereotype about what patients would say and think appeared, in both series of studies, to be so wide of the mark as to lead not only to irrelevant designations, but to erroneous ones. For example, in one paper of Page, Landis, and Katz [10], a 50-item scale for schizophrenia based on expert psychiatric judgment actually yielded *higher* scores for a control sample of nonhospitalized adults than it did for a group of diagnosed schizophrenics. Clearly something is amiss in the clinical conceptions of verbal behavior in psychopathology if errors of this magnitude can appear.

None of the studies mentioned, however, was primarily devoted to specifying and outlining the details of the conventional stereotypes about neuroticism, although all of them revealed something of the grossness of the errors of prevailing opinion. The decision was made, accordingly, to attempt an extensive survey of specific stereotypes about neuroticism and then, by comparing the stereotype with the observed reactions of diagnosed patients, to delineate some of the discrepancies between opinion and fact. In making such a search for discrepancies, the broader the sampling of beliefs and attitudes the better. The Minnesota Multiple Personality Inventory [5], by reason of its systematic coverage of many areas of self definition and personal values, is exceptionally well suited to the task at hand, and was therefore selected as the research instrument.

Method of Analysis

Seven samples of subjects were obtained in order to conduct the analysis. Four of these

[1]This project was conducted under a research grant from the National Institute of Mental Health, National Institutes of Health, U. S. Public Health Service.

Table 1
Sample Items Revealing Erroneous Conceptions about Neuroticism

MMPI item number	Percentage saying "True"						
	Dissemblers			Patients			
	Professional workers ($N = 11$)	Male students ($N = 50$)	Female students ($N = 50$)	Langley Porter males ($N = 39$)	Oakland VA Clinic males ($N = 46$)	Army males ($N = 40$)	Langley Porter females ($N = 51$)
16. I am sure I get a raw deal from life.	64	82	90	8	7	28	12
35. If people had not had it in for me I would have been much more successful.	36	68	86	0	4	2	0
88. I usually feel that life is worth while.	36	28	10	79	74	60	59
212. My people treat me more like a child than a grown-up.	45	76	76	15	20	2	22
257. I usually expect to succeed in things I do.	27	36	10	72	76	88	67
459. I have one or more bad habits which are so strong that it is no use in fighting against them.	73	80	88	8	13	15	12

samples consisted of patients psychiatrically diagnosed as "psychoneurotic," and the remaining three were experimental samples requested to take the role of a patient experiencing a psychoneurotic reaction and to answer the test items accordingly. The samples were constituted as follows:

I. Males
  A. Patients
    1. Thirty-nine patients from the University of California Langley Porter Psychiatric Clinic.[2]
    2. Forty-six patients from the Oakland VA Mental Hygiene Clinic.[3]
    3. Forty patients from a U.S. Army station hospital.[4]
  B. Dissimulators
    1. Fifty students from a University of California, Berkeley, class in abnormal psychology.[5]
    2. Eleven professional workers from the Army hospital.
II. Females
  A. Patients.
    1. Fifty-one patients from the Langley Porter Psychiatric Clinic.
  B. Dissimulators
    1. Fifty students from the class in abnormal psychology.

For each of the items in the MMPI, the proportion in each of the above samples replying "True" was calculated, and then three internal comparisons (I. A. 1 and I. A. 2 vs. I.-B.1; I.A. 3 vs. I.B.2; and II.A. 1 vs. II. B.1) were made. All items yielding similar and statistically significant differences in each comparison were selected. These items, then, could be said to embody the fallacious stereotypes of the experimental samples. The entire statistical table is too long to be given here,[6] but a few selected items are listed in Table 1 to indicate the impressive magnitude of error observed. Note, for example, item number 35: "If people had not had it in for me I would have been much more successful." In the stereotype about

[2]These records were very kindly furnished by Dr. R. E. Harris.

[3]These records were very kindly furnished by Dr. G. S. Welsh.

[4]These records were drawn from an earlier study by the writer [see 1, 2].

[5]These records were very kindly furnished by Dr. T. R. Sarbin.

[6]A complete version of this table, giving all discriminating items and the proportion of each sample answering "True" has been filed with the American Documentation Institute. Order Document No. 3631 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.00 for microfilm or $1.00 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

what patients will say, this sentiment features strongly, and yet, in the actual responses of patients, it is almost nonexistent. In the light of this particular finding what is one to make of a theory which attempts to account for the "fact" that psychoneurotic patients tend to harbor such feelings, and to express them on inquiry?

Altogether, 74 items were identified which met the criterion of significance.[7] These items, and the direction of response given more frequently by the simulators, are listed below:

10. There seems to be a lump in my throat much of the time. (True) 14. I have diarrhea once a month or more. (True) 16. I am sure I get a raw deal from life. (True) 19. When I take a new job, I like to be tipped off on who should be gotten next to. (True) 23. I am troubled by attacks of nausea and vomiting. (True) 24. No one seems to understand me. (True) 29. I am bothered by acid stomach several times a week. (True) 31. I have nightmares every few nights. (True) 35. If people had not had it in for me I would have been much more successful. (True) 42. My family does not like the work I have chosen (or the work I intend to choose for my life work). (True)

44. Much of the time my head seems to hurt all over. (True) 47. Once a week or oftener I feel suddenly hot all over, without apparent cause. (True) 50. My soul sometimes leaves my body. (True) 53. A minister can cure disease by praying and putting his hand on your head. (True) 68. I hardly ever feel pain in the back of the neck. (False) 73. I am an important person. (True) 83. Any man who is able and willing to work hard has a good chance of succeeding. (False) 88. I usually feel that life is worth while. (False) 93. I think most people would lie to get ahead. (True) 96. I have very few quarrels with members of my family. (False)

97. At times I have a strong urge to do something harmful or shocking. (True) 104. I don't seem to care what happens to me. (True) 125. I have a great deal of stomach trouble. (True) 137. I believe that my home life is as pleasant as that of most people I know. (False) 179. I am worried about sex matters. (True) 206. I am very religious (more than most people). (True) 207. I enjoy many different kinds of play and recreation. (False) 210. Everything tastes the same. (True) 211. I can sleep during the day but not at night. (True) 212. My people treat me more like a child than a grown-up. (True)

216. There is very little love and companionship in my family as compared to other homes. (True) 226. Some of my family have habits that bother and annoy me very much. (True) 241. I dream frequently about things that are best kept to myself. (True) 246. My neck spots with red often. (True) 247. I have reason for feeling jealous of one or more members of my family. (True) 257. I usually expect to succeed in things I do. (False) 297. I wish I were not bothered by thoughts about sex. (True) 303. I am so touchy on some subjects that I can't talk about them. (True) 306. I get all the sympathy I should. (False) 320. Many of my dreams are about sex matters. (True)

325. The things that some of my family have done have frightened me. (True) 328. I find it hard to keep my mind on a task or job. (True) 341. At times I hear so well it bothers me. (True) 344. Often I cross the street in order not to meet someone I see. (True) 352. I have been afraid of things or people that I knew could not hurt me. (True) 360. Almost every day something happens to frighten me. (True) 375. When I am feeling very happy and active, someone who is blue or low will spoil it all. (True) 388. I am afraid to be alone in the dark. (True) 405. I have no trouble swallowing. (False) 419. I played hooky from school quite often as a youngster. (True)

422. I have felt embarrassed over the type of work that one or more members of my family have done. (True) 433. I used to have imaginary companions. (True) 438. There are certain people whom I dislike so much that I am inwardly pleased when they are catching it for something they have done. (True) 443. I am apt to pass up something I want to do because others feel that I am not going about it in the right way. (True) 453. When I was a child I didn't care to be a member of a crowd or gang. (True) 458. The man who had most to do with me when I was a child (such as my father, stepfather, etc.) was very strict with me. (True) 459. I have one or more bad habits which are so strong that it is no use in fighting against them. (True) 466. Except by a doctor's orders I never take drugs or sleeping powders. (False) 471. In school my marks in deportment were quite regularly bad. (True) 475. When I am cornered I tell that portion of the truth which is not likely to hurt me. (True)

476. I am a special agent of God. (True) 480. I am often afraid of the dark. (True) 481. I can remember "playing sick" to get out of something. (True) 485. When a man is with a woman he is usually thinking about things related to her sex. (True) 518. I have often felt guilty because I have pretended to feel more sorry about something than I really was. (True) 519. There is something wrong with my sex organs. (True) 524. I am not

[7] The 44 MMPI items revealing the greatest differentiations are included (by permission of the copyright owners) as a "dissimulation scale" in the writer's California Psychological Inventory, copyright 1951. This inventory is one designed to assess variables having broad social and interactional relevance, rather than those pertaining to special medical, psychiatric, or vocational categories. Scales for variables such as social responsibility, dominance and leadership, intellectual efficiency, and academic motivation are included. [See, for example, 3.]

afraid of picking up a disease or germs from door knobs. (False) 525. I am made nervous by certain animals. (True) 528. I blush no more often than others. (False) 535. My mouth feels dry almost all the time. (True)

541. My skin seems to be unusually sensitive to touch. (True) 543. Several times a week I feel as if something dreadful is about to happen. (True) 545. Sometimes I have the same dream over and over. (True) 565. I feel like jumping off when I am on a high place. (True)

The similarity of these items to a conventional neuroticism questionnaire is obvious. Most observers, professionally trained or not, would probably identify the items at once as pertaining to maladjustment. Nevertheless, the inference seems justified that they do *not* pertain to neuroticism and maladjustment, but rather to the prevailing *stereotypes* about neuroticism.[8]

For each of the specific instances listed, neurotic patients answer in precisely the opposite direction from what is predicted for them. They do not, in other words, respond "True" to questionnaire items like "No one seems to understand me," "I don't seem to care what happens to me," and "I have one or more bad habits which are so strong that it is no use in fighting against them," and they do respond "True" to items like "I get all the sympathy I should," and "I usually expect to succeed in things that I do." Even such a short listing as this should serve to highlight the dangers that are involved in any a priori decisions about how patients will answer questions such as these. Indeed, as will be shown below, the way to identify psychoneurotic predispositions as they are manifested on the above list would not be to accept current theory in scoring each item, but rather to reverse it in every instance and score each item backwards.

In order to simplify and condense the list of 74 items, the writer has attempted a subjective clustering. The groupings which were derived are as follows:

1. Numerous physical complaints and dysfunctions.
2. Feelings of victimization, injustice, and misunderstanding; lack of hope for and faith in the future.
3. Exaggerated irritability, petulance, tenseness, fear, and anxiety.
4. Lack of independence and self-sufficiency; inability to face issues squarely.
5. Dissatisfaction with family background and childhood.
6. Sexual conflicts and preoccupations.
7. Bizarre, eccentric ideas; excessive religiosity.

These clusters, then, summarize to some extent the kinds of emphases which are given too much weight in the typical stereotypes about the expressed beliefs and attitudes of maladjusted persons.

Cross-Validating Samples

The separate item discriminations achieved in the analyses were sufficiently striking to suggest trying them as a scale for dissimulation. For this purpose a number of new samples were obtained. The results are indicated in Table 2. Of the 13 samples in Table 2, only one, the 11 Army dissemblers, was used in the item analyses just reported. All of the other samples include cross-validating cases only, so that the results are probably similar to what would be anticipated in most subsequent applications of these items as a scale.

The analysis of variance over 13 separate samples yielded an F ratio of 374.17, significant well beyond the .01 probability point. The clinical samples and the student samples both average near the lower end of the scale, with the dissimulators all scoring much higher.

If the eight clinical samples are pooled in a composite ($N = 915$), a mean of 15.94 and a standard deviation of 9.99 are obtained. For the student samples the combined figures ($N = 507$) are 15.88 and 7.90. The difference between the means is .06, and the CR, 0.125. This is an important finding, for it indicates that patients diagnosed as psychoneurotic actually achieve mean scores on this scale nearly identical with those attained by a large nonhospitalized sample. Patients not only do not respond to the questions in the manner indicated by usual conceptions, but are actually indistinguishable from normals in reacting to the set of 74 statements. Furthermore, both patients and control samples score considerably below the mean achieved by subjects asked to simulate a psychoneurotic reaction.

The distributions of scores for the three con-

[8] The findings also tend to substantiate the viewpoint that structured personality tests may contain subtle and dynamic features over and above what the clinician may recognize in a casual inspection [see 9].

Table 2

Means and Standard Deviations on the 74-Item Dissimulation Scale for the Samples Indicated

Sample	*N*	*M*	*SD*
A. Dissemblers			
1. Professional workers	11	46.82	11.21
2. Students in abnormal psychology	302	53.96	11.52
3. Students in personality in society and culture	41	57.34	11.90
B. Clinical cases			
1. Army psychiatric patients	250	18.87	10.94
2. Ft. Snelling VA Hospital males*	99	16.91	11.45
3. Oakland VA Hospital males**	103	13.40	10.20
4. San Francisco VA Mental Hygiene Clinic males**	62	17.45	9.33
5. Langley Porter Psychiatric Clinic males†	100	15.05	9.21
6. Langley Porter Psychiatric Clinic females†	100	16.25	9.58
7. University of Minnesota Psychiatric Division males‡	101	12.31	7.02
8. University of Minnesota Psychiatric Division females‡	100	13.59	6.86
C. Nonclinical samples			
1. High school freshmen	223	16.89	8.32
2. High school seniors	284	15.08	7.48

* These records were very kindly furnished by Dr. William Hales.
** These records were very kindly furnished by Dr. G. S. Welsh.
† These records were very kindly furnished by Dr. R. E. Harris.
‡ These records were very kindly furnished by Dr. S. R. Hathaway.

solidated samples—dissemblers, clinical cases, and students—are presented in Table 3.

The *F* ratio for the threefold breakdown represented in Table 3 was 2,141.04, a clearly significant value. The screening efficiency of the *Ds* scale for these cross-validation samples (all were cross-validating cases except for the 11 Army dissimulators) can be exampled by a cutting score of 35. Ninety-three per cent of the dissemblers scored at or above this point, whereas only 6 per cent of the clinical cases and 2 per cent of the students did so.

Table 3

Proportion of Cases Falling in the Scoring Categories Indicated

Scores	Dissemblers ($N = 354$)	Clinical cases ($N = 915$)	Students ($N = 507$)
70-74	4.2		
65-69	15.2		
60-64	19.4		
55-59	16.9		
50-54	14.7	0.4	
45-49	11.0	0.8	0.2
40-44	7.3	1.8	0.6
35-39	4.2	2.6	1.2
30-34	3.7	5.3	4.5
25-29	2.0	7.2	7.5
20-24	0.6	12.5	14.8
15-19	0.3	17.5	23.2
10-14	0.3	21.5	24.6
5- 9		19.8	19.3
0- 4		10.1	3.9
M	54.13	15.94	15.88
SD	11.69	9.99	7.90

The results obtained in treating the differentiating items as a *scale* further underscore the erroneousness of the particular aspects of the clinical stereotypes of neuroticism they embody. The opinions and sentiments contained in the items are not expressed, or more accurately, assented to, with any *more* frequency by diagnosed psychoneurotic patients than by control samples, and, furthermore, are agreed with much *less* frequently than the conventional conceptions of neuroticism would presage.

Summary

An investigation of typical conceptions of neuroticism was conducted by having subjects take the role of a psychoneurotic patient in responding to a personality inventory. Significant discrepancies between what diagnosed patients did, in fact, report on this inventory and the stereotypes given by the simulators were discovered.

The argument was advanced that errors of the magnitude observed suggested a considerable degree of misinformation in the prevailing conceptions about neuroticism.

Finally, it was possible, by consolidating the items individually tapping these misconceptions, to develop a measuring instrument yielding similar scores for both patient and control samples, and differentiating each from samples based on dissembled records.

Received November 16, 1953.

References

1. Gough, H. G. Simulated patterns on the MMPI. *J. abnorm. soc. Psychol.*, 1947, **42**, 215-225.
2. Gough, H. G. The *F* minus *K* dissimulation index on the MMPI. *J. consult. Psychol.*, 1950, **14**, 408-413.
3. Gough, H. G. The construction of a personality scale to predict academic achievement. *J. appl. Psychol.*, 1953, **37**, 361-366.
4. Hathaway, S. R., & McKinley, J. C. A Multiphasic personality schedule: I. Construction of the schedule. *J. Psychol.*, 1940, **10**, 249-254.
5. Hathaway, S. R., & McKinley, J. C. *The Minnesota Multiphasic Personality Inventory.* Minneapolis: Univer. of Minnesota Press, 1943.
6. Landis, C., & Katz, S. E. The validity of certain questions which purport to measure neurotic tendencies. *J. appl. Psychol.*, 1934, **18**, 343-356.
7. Landis, C., Zubin, J., & Katz, S. E. Empirical evaluation of three personality adjustment inventories. *J. educ. Psychol.*, 1935, **26**, 321-330.
8. McKinley, J. C., & Hathaway, S. R. A multiphasic personality inventory: V. Hysteria, hypomania, and psychopathic deviate. *J. appl. Psychol.*, 1944, **28**, 153-174.
9. Meehl, P. E. The dynamics of structured personality tests. *J. clin. Psychol.*, 1945, **1**, 296-303.
10. Page, J., Landis, C., & Katz, S. E. Schizophrenic traits in the functional psychoses and in normal individuals. *Amer. J. Psychiat.*, 1934, **90**, 1213-1225.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Factors Descriptive of Chronic Schizophrenics Selected for the Operation of Prefrontal Lobotomy

Maurice Lorr, Richard L. Jenkins, and James Q. Holsopple
Veterans Administration, Washington, D. C.

This study was undertaken in the hope of contributing to a simpler and conceptually more satisfactory differentiation of the major schizophrenic processes. To this end a set of scores descriptive of a group of chronic schizophrenics, chosen for the operation of prefrontal lobotomy, and their matched controls were subjected to a multiple-factor analysis.

Procedure

The data consisted of clinical judgments recorded on the Northport Record [4, 7], a schedule of 81 brief rating scales for describing the characteristic behavior and symptoms of psychotic patients. The ratings were made by psychiatrists and psychologists on 153 schizophrenic patients at six Veterans Administration hospitals which are participating in a study to define the nature and extent of changes that occur as a result of the operation of prefrontal lobotomy [3]. All ratings were obtained prior to operation on patients later lobotomized or on patients selected to serve as controls. Since no patient had been mentally ill for less than 4 years, the data represent a group of patients whose psychoses had become chronic.

The 11 scores included in the study were based on a factor analysis of ratings of 423 hospitalized patients on 55 of the scale variables included in the Northport Record [5]. Each score consisted of the sum of the ratings received by the patient on a group of scales best defining a factor. For example, the paranoid projection factor *C* consisted of the sum of the ratings on scales descriptive of ideas of reference, delusions of persecution, delusions of influence, and frequency of assaults. There were relatively few scales common to two or more factors. Factor *A*, depressed retarded versus manic excited, and factor *E*, melancholy agitation, had 4 scales in common. A few of the factors had two scales in common. Of the 55 possible relations between factors there is no overlap in 34.

The 11 factors scored have been named as follows:

A. Retarded depression versus manic excitement

B. Compliance versus resistiveness

C. Paranoid projection

D. Activity level

E. Melancholy agitation

F. Perceptual distortion

G. Motor disturbances

H. Submissiveness versus belligerence

I. Aggressive social interest versus social withdrawal

J. Self-depreciation versus grandiose expansiveness

K. Conceptual disorganization

The Pearson correlations computed between the 11 factor scores are presented in Table 1. The correlational matrix was subjected to a centroid analysis with centroid estimates of the communalities in the diagonals, and 3 factors were extracted. The median absolute third factor residual was .06 as compared to the standard error of .08 of a zero correlation on a population of this size. The orthogonal factors and the rotated oblique factors are presented in Table 2. The cosines between the reference vectors are as follows: Alpha versus Beta, .07; Alpha versus Gamma, .51; Beta versus Gamma, .05. The transformation matrix is shown in Table 3.

Table 1
Intercorrelations of Raw Scores on 11 Factors of 153 Schizophrenic Patients

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
A	−.13									
B	.01	.03								
C	.16	.02	.02							
D	.53	−.28	.06	.04						
E	−.04	.39	.46	.10	−.22					
F	−.11	.42	.17	.06	−.27	.02				
G	.32	.50	.33	.23	.25	.16	−.24			
H	−.54	.39	−.06	−.23	−.36	.01	.35	.71		
I	.29	.40	.14	.22	−.05	.27	.12	.57	−.01	
J	−.28	.53	.28	.00	−.47	.71	.63	−.08	.40	.11

The Findings

The first factor, Alpha, is defined by withdrawal (*I*), conceptual disorganization (*K*), motor disturbances (*G*), absence of melancholy agitation (*E*), the schizoid elements of depressive retardation (*A*), and resistiveness (*B*). A patient with high scores on Alpha exhibits manneristic movements and postures, lacks tension or anxiety, appears apathetic or lacking in feeling. Typically he is mute or taciturn, and when he responds his speech tends to be irrelevant, incoherent, and loaded with stereotyped language. He is also likely to be disoriented for self and others. On the ward he will sit isolated, mute, and indifferent to the activities around him. The parameter underlying Alpha appears to be one of apathetic withdrawal with motor disturbances and conceptual disorganization.

The second-order factor, Gamma, which is negatively related to Alpha, is most sharply defined by factor *H* (submissiveness versus belligerence), factor *I* (withdrawal), factor *B* (compliance versus resistiveness), and factor *J* (self-depreciatory versus grandiosity). At one end Gamma is characterized by overt hostility, irritability, assaultiveness, obscenity and profanity, neglect of personal appearance, isolation from others, resistiveness, and destructiveness. Ideation is marked by ideas of grandeur and boastful conceit. The opposite end of Gamma is characterized by a meek submissiveness, compliance to routine, and self-depreciatory tendencies.

The belligerent, resistive, assaultive behavior of Gamma is often associated with an excited agitation. This would account for the substantial negative relation between Alpha and Gamma although the two also have in common withdrawal and resistive tendencies. It is the type of patient high on Gamma who is frequently a candidate for lobotomy.

Table 2
Second-Order Centroid Matrix F and Oblique Matrix V

Factors		F				V		
		I	II	III	h^2j	α	β	γ
Depression*	*A*	—27	61	18	48	55	11	—07
Resistance	*B*	70	07	27	57	49	36	66
Paranoid distortion	*C*	34	25	—24	23	—02	47	08
Overactivity	*D*	09	30	—07	10	—17	20	06
Apathy*	*E*	—45	50	16	48	60	27	08
Perceptual distortion	*F*	60	26	—55	73	05	85	01
Motor disturbances	*G*	49	—36	09	37	60	18	31
Belligerence	*H*	44	63	62	98	04	14	86
Withdrawal	*I*	61	—39	60	89	85	—06	77
Grandiosity	*J*	42	53	16	48	—05	37	48
Conceptual disorganization	*K*	78	—30	—29	79	64	65	22

* A and E and their descriptions have been reversed in matrix V.

Table 3
Transformation Matrix

	A	*B*	*C*
I	.66	.72	.63
II	—.69	.32	.18
III	.29	—.62	.75

The second-order factor, Beta, has its highest correlations from perceptual distortion (*F*), conceptual disorganization (*K*), and paranoid projection (*C*). A patient scoring high on this factor is characterized by auditory and visual hallucinations, ideas of reference, and delusions of persecution and influence. He is likely to be irrelevant or incoherent in his speech, and disoriented as to who he is or who

others are around him.

Schizophrenic patients considered for lobotomy may, in the foregoing terms, be said to show three distinguishable modes of reaction. They withdraw (Alpha), they distort in perception and thought the world they encounter around them (Beta), and /or they fight (Gamma). If the withdrawal factor Alpha is pronounced and the disorganization of thinking factor Beta is not obvious, the patient is likely to be diagnosed as showing catatonic withdrawal. If disorganization (Beta) is pronounced, the patient will tend to be diagnosed as hebephrenic. If the progress of the disorganization is slowed by the false solution of paranoid projection, the patient will be categorized as a paranoid schizophrenic. If an assaultive belligerence (Gamma) is marked and associated with an agitation (negative Alpha) the patient will be regarded as a good candidate for lobotomy.

Discussion

The three second-order factors identified here will now be compared with those identified by Degan [1], and by Lorr, Jenkins, and O'Connor [5]. Since the patient groups sampled differed considerably in these studies and the scales used were different, we can expect only a general correspondence in the factors identified. Nevertheless, it is of interest to see how much these second-order factors have in common.

Degan utilized data collected by Father T. V. Moore [6] on 367 hospitalized patients diagnosed as functional psychotics. He identified nine first-order factors and four in the second-order domain. Table 4 presents the projections of his first-order factors on the second order. In the study by Lorr *et al.* [5], 55 ratings descriptive of 423 hospitalized psychotic patients were factored. The 11 first-order factors were analyzed and three second-order factors were identified (Table 5).

Table 4

Degan's Second-Order Rotated Oblique Matrix

First-order factors		Second-order factors *W*	*X*	*Y*	*Z*
Hyperprojection	*A*	–09	40	–52	–01
Depression	*B*	–06	–13	51	–42
Manic hyperexcitability	*C*	56	13	–03	00
Catatonia	*D*	–08	–03	–01	71
Schizophrenic dissociation	*E*	–01	68	01	03
Traumatic hysteria	*F*	01	45	–20	35
Hyperirritability	*G*	61	–01	–37	33
Deterioration	*H*	22	19	17	32
Neurasthenia	*I*	–26	10	63	–03

Degan's second-order factor *Z*, interpreted as catatonic schizophrenia, is most similar to factor Alpha. Factor *Z*, in terms of the original variables, is defined by mutism, refusal of food, stereotypism of attitude, negativism, giggling, and absence of an agitated depression. To a lesser degree, it is characterized by disorientation for time and space, irritability, tantrums, and loss of finer sensibilities. Only the irritability elements of factor Z are lacking in our Alpha. It should be noted that Alpha was identified in a group of patients diagnosed as schizophrenic and sufficiently chronic to be considered for lobotomy, while factor *Z* was derived from a mixed psychotic population. Furthermore, in the population from which Alpha was drawn, assaultiveness, hostility, and irritability were sufficiently common as to define factor Gamma (belligerence). It is for these reasons that irritability has its projections on Gamma rather than Alpha.

Table 5

Second-Order Oblique Factor Matrix Based on Intercorrelations between Primaries (Lorr, Jenkins, O'Connor [5])

Factors		*AA*	*BB*	*CC*
Depression	*A*	47	25	34
Resistiveness	*B*	54	06	22
Paranoid distortion	*C*	–08	39	–03
Overactivity	*D*	–21	–42	01
Apathy	*E*	–03	00	54
Perceptual distortion	*F*	09	34	02
Motor disturbances	*G*	56	01	00
Belligerence	*H*	58	21	03
Withdrawal	*I*	45	04	68
Grandiose expansiveness	*J*	–20	–10	26
Conceptual disorganization	*K*	–29	03	13

Factor *AA* (Table 5) is characterized by resistiveness, frequent assaults, manneristic movements and postures, disorientation for time, place, self, and people, a complete lack of interest, and muteness. Alpha has in common with *AA* manneristic movements and postures, disorientation, and withdrawal of interest, but it lacks the assaultiveness of *AA*. In addition, Alpha is characterized by a conceptual disorganization missing in *AA*.

Beta has considerable resemblance to Degan's factor *X*, which is defined principally by stereotypism of words and of actions, talking to voices, appearing "shut-in," disorientation for space and time, hallucination, and bizarre delusions. Only the resistiveness and grandiosity of Beta are absent.

Beta is probably the same as factor *BB* in Table 5. Factor *BB* is defined by delusions of influence, ideas of reference, ideas of persecution, auditory and visual hallucinations, and slowed activity. It lacks only the thinking disorganization of Beta.

Since the present study was confined to severe chronic schizophrenics, it is not surprising that primarily affective factors present in the other two studies are missing. In Degan's study, we have a factor of manic excitement (*W*) and a bipolar factor of paranoid versus depressive psychosis (*Y*), i.e., of projections versus introjection. Factor *CC*, on the other hand, is based on an analysis of a population probably having fewer patients with manic-depressive psychosis and more with schizophrenia. As a consequence, factor *CC* is more schizophrenic than affective. It ranges from an apathetic withdrawal at one end to excitement with grandiose expansiveness at the other. Factors *W*, *Y*, and *CC* do not appear in the present study of severe chronic schizophrenics.

The belligerence factor Gamma is unique to the present study. No doubt it is due to the selection of a sample of patients considered for lobotomy, many of whom were explosively tense and assaultive. Resemblances between the factors are summarized in Table 6.

Table 6

Similarities between Second-Order Factors Identified in Three Factor Analyses

Hypothesized process	Study: Degan's analysis	Study: Lorr, Jenkins, O'Connor analysis	Present analysis
Resistive withdrawal	*Z*	*AA*	Alpha
Disorganization and distortion of thinking	*X*	*BB*	Beta
Manic excitement	*W*	*CC*(?)	Not found (schizophrenic patients)
Projection versus introjection	*Y*		
Belligerence	Not found	Not found	Gamma

Summary

A factorial analysis was made of the intercorrelations between 11 relatively independent scores descriptive of the behavior and psychopathology of 153 chronically ill schizophrenic patients selected for prefrontal lobotomy and their matched controls. Three common factors were extracted and rotated to simple structure. Alpha is interpreted as a factor of apathetic withdrawal with motor disturbances. Beta is a factor of perceptual and thinking distortion which may represent a phase of schizophrenic disorganization. Gamma describes a fighting reaction. The negative relation between Alpha and Gamma suggests a third-order bipolar parameter of apathetic withdrawal versus agitated belligerence.

The first two factors resemble those found in Degan's analysis of Father Moore's data and those isolated in the Lorr, Jenkins, and O'Connor study.

Received November 18, 1953.

References

1. Degan, J. W. Dimensions of functional psychosis. *Psychometr. Monogr.*, 1952, No. 6.
2. Jenkins, R. L. The schizophrenic sequence: withdrawal, disorganization, psychotic reorganization. *Amer. J. Orthopsychiat.*, 1952, **22**, 738-748.
3. Jenkins, R. L., & Holsopple, J. Q. Criteria and experimental design for evaluating results of lobotomy. *Res. Publ. Ass. nerv. ment. Dis.*, 1951, **31**, 319-327.
4. Lorr, M. Multidimensional scales for rating psychiatric patients. Hospital form. *Vet. Adm. tech. Bull.* 10-507, Nov. 16, 1953.
5. Lorr, M., Jenkins, R. L., & O'Connor, P. Factors d scriptive of psychopathology and behavior of hospitalized patients. *J. abnorm. soc. Psychol.*, in press.
6. Moore, T. V. The essential psychoses and their functional syndromes. *Stud. Psychol. Psychiat. Cathol. Univer. Amer.*, 1933, **3**, No. 3.
7. U. S. Veterans Administration. Development of the Northport Record. *Infor. Bull.* 10-5-21, Psychiat. Neurol. Issue, July, 1950.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Group Therapy Techniques with Hospitalized Chronic Schizophrenic Patients

Joseph M. Sacks and Stanley Berger
Franklin D. Roosevelt VA Hospital, Montrose, N.Y.[1, 2]

Until recent years, the predominant approach to group treatment of hospitalized schizophrenic patients called for mass activities or didactic lectures [4, 5, 7, 8, 11]. More recently, there has been an increased application of psychodynamic principles and techniques in group psychotherapy with psychotics [1, 3, 6]. Most dynamic approaches to group psychotherapy have in common the expression and release of feelings, mutual understanding and emotional support, recognition of the individual's highly personalized needs, an opportunity for each member of the group to test his behavior, thinking, and feeling against the reality of his impact upon others, a sharing of common problems with others, and the cooperative working out of solutions to these problems with ultimate self-understanding and acceptance. The authors of this paper believe that the group treatment of psychiatric patients involves all of these elements and something in addition which is of crucial importance: the understanding and exploration of distortions in the perceptions and behavior of the patients in their interpersonal relationships. This paper presents a description of the techniques employed and results obtained in an attempt to apply psychodynamic principles to the group treatment of chronic schizophrenics.

The techniques to be described have been applied for the past year to four groups of 28 patients on the Continued Treatment Service of the Franklin D. Roosevelt VA Hospital. The groups were conducted by cotherapists, including a psychiatrist, a staff psychologist, and a clinical psychologist trainee. The groups met twice weekly for periods of an hour and a half.

A control group was derived from the hospital admissions records of patients admitted to the Continued Treatment Service on the same day as the patients of the therapy group. No significant differences were found between the two groups with respect to age, socioeconomic level, education, or intelligence.

Theoretical Considerations

In our approach to group therapy, we incorporated some of the more basic theoretical formulations that have been fruitful in treating less severely disturbed patients. In general, the concepts discussed above were integrated into a philosophy of therapy which we felt would be of value in any group psychotherapy. However, the focusing of treatment upon chronic, regressed schizophrenic patients necessitated a broadening and elaboration of some of the more succinctly presented theoretical concepts, as well as additional formulations peculiar to our patient group.

We have been guided by the following assumptions:

1. One of the primary characteristics of schizophrenic patients, according to certain workers [9, 10, 12], is their acute sense of deprivation; that is, they react as though they have experienced the trauma of never having had, or of losing, a mother-figure who normally would have satisfied their physical needs (often symbolized by food), as well as their emotional needs, by providing comfort and consistent love, protection, and security. In reaction to these basic deprivations with their inherent threat to life itself, the infant has been forced

[1] Submitted with the approval of the Chief Medical Officer, Veterans Administration, who assumes no responsibility for the opinions expressed here, which are those of the authors.

[2] The authors gratefully acknowledge the participation and supervision of the following ward psychiatrists: Leopold Jordani, Maurice Pachter, and Paul Wenger. We also wish to thank Edwin Barker, Hanna Husserl, Warren Fox, Nathaniel Wagner, and Helen Singer, clinical psychologist trainees, who participated in various phases of this project.

to draw upon modes of adjustment and perception which are highly distorted and which, persisting to adulthood, prevent the development of more healthy and satisfying psychological mechanisms of defense.

2. The reliance upon inappropriate and distorted defenses is experienced by the patient as highly inadequate, fraught with potential danger and anguish, and in constant threat of collapse. Federn [2] stresses the point that the schizophrenic patient is constantly defending himself against becoming aware of the original fear of isolation and deprivation (death), and must continually apply whatever defenses he can to manage some sort of psychic equilibrium or reduction of tension. However, the patient is aware (at some level of consciousness) that these defenses are not effective.

3. Many techniques applicable to neurotic groups, or even to groups of ambulatory schizophrenics or schizophrenics in remission, are not effective in treating chronic schizophrenic patients. Collectively, these patients lack sufficient integration or group cohesion so that techniques making use of appeals to reason or judgment ordinarily fail. In the same manner, attempts to utilize the patients' ability to test reality or to control their impulses rarely meet with therapeutic benefit.

4. The most immediate task in group therapy with chronic schizophrenics is to build bridges of communication among members of the group and between patients and therapists. The inability to establish and maintain healthy and satisfying interpersonal relationships makes the schizophrenic patient extremely difficult to treat. Techniques that permit the therapist to reach the patient should be used and nurtured in the belief that they provide healthy avenues to the personality; this gives the therapist a valuable, though delicate, hold upon the inner desperate hope for health which, it is assumed, almost every patient has.

We believe that the techniques described below are valuable in a chronic-schizophrenic group setting, and that their application provides a successful means of treating these patients. In the following section we shall enumerate briefly the specific techniques and the theoretical rationale that have been employed in connection with various important components of the therapeutic situation. When possible, specific examples drawn from tape-recorded transcriptions of group sessions are included.

Technique

The therapy. Making contact with the patients requires a very active role on the part of the therapist, for contact cannot be accomplished by a passive-permissive attitude no matter how kindly or well intentioned. Silence, for example, begets more silence. We have seen a therapy group sitting together for an hour without uttering a single word. These patients, often withdrawn into their own autistic fantasies and reveries, cannot be brought into direct contact with each other without active stimulation, most frequently by a direct discussion of the silence. The following transcription demonstrates the active role of the therapist in building up a positive relationship in a patient who has exhibited marked hostility and assaultiveness:

P. You make it sound as if my father didn't give a damn about me.
T. That's how you feel?
P. My grandfather was someone I could get along with much better. He was the kind of fellow who gave advice and respected my opinions. He's a man who would shake your hand.
T. I will shake your hand.
P. You would?
T. Yes.
P. You mean as a sign of respect?
T. Yes.
P. Well then, here.
Patient extends hand; therapist walks over to patient and shakes his hand. Patient bursts into tears.

Another fruitful technique for active stimulation is to facilitate the patient's understanding of the deprivations which he experienced early in life and the manifestations of such needs in the present situation. Material which has meaning on other levels of adjustment is most frequently dealt with on this early level, assuming that this is most fruitful and meaningful for these patients [9]. If less direct methods are employed, patients respond with evasive, confused, autistic material, often engaging in long meaningless monologues. When the patients are able to resolve some of the difficulties in their earliest relationships, increasingly greater use is made of material related to later per-

iods of adjustment. However, the early deprivations have been so severe that there is a constant return to problems arising out of this period. The handling of these feelings of infantile deprivation, in which the underlying fear of death is revealed, is demonstrated by the following example:

P. I would be all right if it weren't for the pain.

T. What pain?

P. It grips my throat, it hits me here [cheek] and my teeth lock together.

T. When does this happen?

P. It's from the whisper.

T. What whisper?

P. They say those nasty things. I have to fight them. I'd be all right if I didn't have to eat that medicine.

T. What medicine?

P. The stuff they put in my food. It makes my body jolt. I get these awful pains.

T. Why do you eat the medicine?

P. I have to. It keeps me alive. The doctors made my mother give it to me when I was a child so I wouldn't end up in a hospital.

T. You're in a hospital now.

P. Yeah, but I'm not dead.

T. You say you were afraid that you would die if you didn't eat the bitter medicine your mother gave you?

P. You're not kidding, Doc!

In order to validate our understanding of the patient's acute sense of oral deprivation in terms of the theoretical considerations discussed above, and our willingness to try to satisfy them, we bring "food" (chewing gum and chocolate) to the group sessions. It is not merely the direct physical satisfaction of oral cravings that is involved here, but more significantly, the process of receiving food from the parental figures as represented by the therapist, and partaking of it in the family setting of the group. The patients respond quite favorably, and it is noteworthy that in some cases the patients talk about buying candy for the therapist, or offer him cigarettes during the session. At one session a patient brought in a bun from his breakfast and placed it on the table in the center of the group for all to share.

In a further attempt to construct some relationship with the patients, the therapist may resort to overt attempts to dress in a similar fashion (no tie, sleeves rolled up), sit in a similar manner (slouched, or at times stiffly), and speak in a similar manner. In terms of the latter, the use of common language is certainly more meaningful to these patients, and at the same time is, in a sense, granting them permission to speak freely and honestly. The following example points out the importance of communication with the patient in his own language as a means of building up positive relationship:

P. I don't know how many personalities I have.

T. Why don't you know?

P. Because I always thought a person could talk to another person in an understandable way. That is personality.

T. Can you talk to a person in an understandable way?

P. I mean if he is angry, or if he is sad. Personality to me seems like I feel when I have medicine; the way you talk to a person, the way you answer.

T. It must have something to do with the way you talk to me. How do you talk to me?

P. I mean you talk not only because your words are well spoken, or the sound of your voice, but you can understand what I mean, and I can understand what you mean.

T. We understand each other.

P. Sort of a love affair. That's what I think personality is.

T. You love me because I understand you?

P. That's right.

The therapist. A difficult task for the therapist is to permit a relaxed and receptive atmosphere within himself so that he can be more sensitive to the highly symbolic associations of the patients, and can rely upon his own clinically intuitive responses to aid in communicating with them. To maintain this sensitivity to communication requires a healthy personality for the therapist, who can, by accepting his own impulses and needs in a mature manner, be more accepting of the patient's desires and impulses. It has been often observed that schizophrenic patients are very perceptive, and they will be quick to observe in the therapist anxiety and fear, and especially rejection.

The therapist maintains himself as the central figure in the group about which therapy pivots. Though the patients may shift about, the therapist always sits in the same seat, often indicating the strength and valence of the relationships, the patient's need for contact, and his desire for communication.

The consistency of the therapist is further emphasized by his remaining with the group despite other hospital assignments or commitments. If the group is transferred to another

building, it is desirable for the therapist to be transferred likewise. Such maintenance of group integrity provides a group-stabilizing force, as well as a means of helping the patients make these kinds of moves and new adjustments.

The therapist continually attempts to build up positive feelings and tries to deal with the negative ones. Hostility, rivalry, and competition within the group must not be allowed to continue too long. When too much hostility is expressed by one patient to another it must be redirected (preferably to the therapist) and reduced in strength so that the angry patient and the rest of the group are not overwhelmed by an overt display. Prolonged uncontrolled expression of strong feeling is potentially upsetting to schizophrenic patients who are in constant dread of losing control over themselves, and it is possible for one agitated patient to set off the entire group if the therapist fails to be the strong, controlling figure. The following transcription illustrates bringing patient to patient hostility out in the open and redirecting it toward the therapist:

A. [For about fifteen minutes has been playing a didactic therapist role toward another patient, B, newly admitted to the group, asking questions about his history, his plans for the future, and finally giving a case summary.] So it looks as though you want to depend on Uncle Sam for the rest of your life. Don't you see that this dependency is part of your sickness?

C. [Has been squirming in his seat, flushing, left the room and returned.]

T. [to C] How have you been feeling these last few minutes?

C. Like I'm going to explode.

T. Why is that?

C. Listening to A shoot off his mouth. He isn't trying to help B, he is just trying to show how much he knows.

A. That's not true. We are supposed to help each other here, aren't we doctor?

T. Maybe you're doing to B what you feel I have done to you in the past, pointing out your own dependency needs. You must have felt pretty angry when I did that.

B. May I say something? Yes, I did resent what A was doing but also I felt he was trying to help me and I appreciate it.

C. He wasn't trying to help. He was just trying to play the doctor.

A. Maybe you've got something there, C. I certainly wasn't aware of it though. I guess you do resent it, doctor, when someone points out your own weaknesses to you.

[After the session A and C go out together talking with much animation, but in friendly fashion rather than in a hostile manner.]

The group. The patient's illness is thought of in terms of "ego-sickness" [2]. When patients request an answer to some realistic question, or when it is felt beneficial to discuss some aspect of the schizophrenic process, the explanation is centered about the patient's inability to reconcile bodily impulses with an unrelenting superego, and the resulting disintegration or fragmentation of the personality structure. When the therapist feels that it is beneficial to reply to a question by a patient concerning his illness, the reply should be honest and straightforward. Such frankness quickly earns the patient's respect. Dealing with the patient's illness openly and honestly is reassuring, not threatening as illustrated below:

P. So my father told me to come up here for a couple of weeks to rest and to gain some weight. Well, I've had enough rest, I've been in this hospital for more than a year. But they don't want me to go home yet. I don't understand it.

T. The truth is that you have an emotional disturbance. That's why you are in this hospital. This is no rest home.

P. Oh, I guess my father knew what he was doing.

T. What do you mean?

P. I couldn't control my emotions when I was home. I was blowing my top all the time, couldn't keep a job, afraid of people. I had to come up here to straighten out all that.

T. What about not understanding why you can't get out yet?

P. I guess that will have to wait until I straighten myself out a lot more.

The administrative problems of ward and hospital, as they affect the patient, cannot in themselves be considered a part of group therapy. It should be quickly pointed out to the patients that the purpose of the therapy is not to grant passes or leaves of absence, assign patients to work details, or engage in other administrative duties of this nature. These factors are certainly considered by the therapists in terms of what arises in therapy. However, the therapist reinforces his position in the group as a person sincerely interested in helping the patients get better, rather than one who essentially grants or denies privileges.

The solidarity of the group is emphasized by assigning the patients to the same sleeping and eating quarters, the same work details, and in general keeping them together on many hospital activities. While these procedures may involve considerable administrative difficulty, they have been found to be therapeutically beneficial.

The making of tape recordings and playing them at the subsequent session has been found to be a useful technique for enabling the group to gain a more objective and less distorted view of themselves. A frequent reaction is surprise at the amount of feeling displayed about a given subject. Schizophrenic patients in particular seem to derive a great deal of satisfaction from hearing their own voices and are eager to listen. They will interrupt the playing of the recording to elaborate something which they have said, or to speak of other aspects of the discussion. At the end of playing one such recording, dominated largely by two patients, one of these two turned to the other and said, "Let's you and me take a back seat next time and give the other fellows a chance." The following shows how the review of a previous session by tape recording enables a patient to interpret his own symbolic utterance and to present new material for further therapeutic exploration.

P. It felt like I was choking, like there was a rope around my neck. Maybe one of my ancestors was a pirate, no, maybe I was a pirate in a previous life and I was hung by the neck from the yardarm of a ship.

T. Why were you being hung?

P. I don't know. I never committed any crime. [Patient interrupts recording at this point.] The only crimes I have ever committed were against myself. I was thinking about how I threw my life away. That choking feeling has nothing to do with pirates. It was my own disgust with myself and the thought of suicide coming out in the form of a symptom. That's what it was.

Individual conferences with the therapist are available to the patients when they so desire or when the therapist feels that a patient might return to the ward in an agitated state. It was in one of these individual sessions that a patient confided, for the first time, that he had suicidal ideas. Others have been able to talk about fantasies or impulses which they had previously suppressed.

Results

In our initial meetings with these patients, on the more inactive wards in the hospital, many of them were very withdrawn, apathetic, autistic, uninterested in the environment, and unaware of their own lack of personal tidiness and crude behavior. In no case did the patients appear to have any understanding of, or any interest in, group therapy. The first few sessions in the group found them to be uncommunicative and withdrawn, they showed no interest in the other patients or the therapist, and mainly sat for the entire time lost in their own preoccupations.

At the present time we see most significant changes in the patients' interpersonal experiences and behavior. The group meetings are active and alive, often with several patients trying to talk at the same time. They are interested in themselves and their environment, and even more noteworthy, they have demonstrated an increasing interest in each other and in the therapist. In addition, many of the patients have become interested and engaged in hospital activities and work details. In one case, an entire group was transferred to a rather active ward on which they are all assigned to work details, engage in athletics, and regularly attend hospital parties and movies.

These patients have made definite advances in achieving better emotional adjustment, and they are highly motivated for and better capable of receiving intensive psychotherapy. We feel that the therapeutic techniques employed have enabled us to penetrate long-standing psychic barriers and to initiate more healthy and gratifying modes of adjustment in pa-

Table 1
Hospital Status of Psychotherapy Group and Control Group After One Year

Status	Control group	Therapy group	D_σ	p
On More Disturbed Ward	4 (14%)	6 (21%)	.80	.42
Unchanged	18 (64%)	5 (18%)	3.53	.0004
On Improved Ward*	1 (4%)	13 (46%)	3.39	.0008
Discharged or Trial Visit	5 (18%)	4 (14%)	.37	.70
Total	28	28		

* Wards requiring better adjustment to other patients and personnel, and more active participation in the hospital program.

tients who might have otherwise remained lifelong, regressed schizophrenics.

It is apparent from Table 1 that there are no significant differences between the therapy and control groups in the "More Disturbed" and "Discharged or Trial Visit" categories. On the other hand, in the "Unchanged" and "Improved Ward" categories, the differences in favor of the therapy group are very large. In the control group, 64% are unchanged, compared with 18% of the therapy group. This difference is significant at the .0004 level. In the therapy group, 64% are improved, compared with 4% of the control group. This difference is significant at the .0008 level.

When the status categories were combined into Improved (On Improved Ward plus Discharged or TV) and Unimproved (More Disturbed Ward plus Unchanged), a chi square of 9.02 was found, indicating a probability of less than one in a hundred that the differences between the therapy and control groups are due to chance. These results are in the same direction as those obtained by Powdermaker and Frank [6, ch. XVII] and by Feifel and Schwartz [3].

Summary

This paper describes, in the opinion of the authors, some of the more effective group therapy techniques, which were actively applied for a period of one year to four groups of chronic schizophrenic patients in a large hospital setting. A clinical evaluation of the patients' improvement is offered, as well as statistical comparison with a control group.

Theoretical assumptions upon which these techniques have been based are presented. The techniques themselves are discussed in terms of the therapist, the group, and the therapy process.

Illustrations from recorded transcriptions of group therapy sessions are given as concrete illustrations of several of the points presented.

It is hoped that the presentation of this material will stimulate greater interest in the chronic schizophrenic patient, and that continued efforts will eventually discover therapeutic techniques that will prove fruitful and efficacious in restoring these patients to a more satisfactory way of life.

Received November 20, 1953.

References

1. Burchard, E. M. L., Michaels, J. J., & Kotkov, B. Criteria for the evaluation of group therapy. *Psychosom. Med.*, 1948, **10**, 257-274.
2. Federn, P. *Ego psychology and the psychoses.* New York: Basic Books, 1952.
3. Feifel, H., & Schwartz, A. D. Group psychotherapy with acutely disturbed psychotic patients. *J. consult. Psychol.*, 1953, **17**, 113-121.
4. Klapman, J. W. Group psychotherapy in institutions. *Group Psychother.*, 1951, **4**, 181-192.
5. Marsh, L. C. Group treatment of psychoses by psychological equivalent of the revival. *Ment. Hyg., N. Y.*, 1931, **15**, 328-349.
6. Powdermaker, Florence B., & Frank, J. D. *Group psychotherapy.* Cambridge, Mass.: Harvard Univer. Press, 1953.
7. Pratt, J. H. Results obtained in the treatment of pulmonary tuberculosis by the class method. *Brit. med. J.*, 1908, **2**, 1070-1079.
8. Pratt, J. H. The principles of class treatment and their application to various chronic diseases. *Hosp. soc. Serv.*, 1922, **6**, 401-411.
9. Rosen, J. *Direct analysis.* New York: Grune & Stratton, 1953.
10. Sechehaye, M. A. *Symbolic realization.* New York: International Universities Press, 1951.
11. Slavson, S. R. *Analytic group psychotherapy.* New York: Columbia Univer. Press, 1950.
12. Sullivan, H. S. *Conceptions of modern psychiatry.* Washington: W. A. White Psychiatric Foundation, 1947.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

The Influence of Fullness of Interview on the Reliability, Discriminability, and Validity of Interview Judgments[1]

Harry Levin
Harvard Graduate School of Education

Every interviewer has had the experience of talking both with laconic and with verbose respondents. For clinical purposes, the length of response may in itself be a significant factor. However, when interviews are used to gather research materials—that is, when the responses will be quantified and related to other variables—we may ask how the fullness of the responses affects the judgments made on the interview, assuming that both the short and the long replies are codable.

Factors determining the length of replies have not been widely examined. Both Shapiro and Eberhardt [4] and Feldman, Hyman, and Hart [1] found that there was wide variability among interviewers as to how much they probed the respondent. It is likely that the more extensive the probing, the longer the interview. Doubtless the general verbality of the respondent also influences the length.

Given interview responses of different lengths, or fullness, however, one may ask what influence this factor has on the reliability, discriminability, and validity of the coded judgments made from the material. A set of 379 transcribed interviews obtained in connection with a study of maternal child-rearing practices has offered an opportunity to make preliminary tests of three hypotheses.

Three kinds of codes were used to quantify the interviews: (*a*) item codes (henceforth labeled I) required judgments on responses to single specific questions; (*b*) summary codes (S) involved judgments based on a group of three or four questions; (*c*) over-all codes (O) took into account material which occurred not only in response to specific questions, but also at various points throughout the interview.

The following hypotheses were tested:

1. Reliability on all three types of codes is higher for the short than for the long interviews. The reasoning is that the added materials lead to different interpretations by the coders.

2. Discriminability of judgments is greater for the long than for the short interview. This prediction is based on the idea that more material in the interview permits finer discriminations among respondents.

3. Long interviews lead to more valid judgments than short interviews. It seems likely that added materials lead to judgments which are "truer" to behavior than are those based on sparser materials.

Procedure

Three hundred and seventy-nine mothers were interviewed about various aspects of their child-training procedures. A standardized interview schedule was used, consisting of a list of open questions which the interviewers were trained to use without change in wording and in the prescribed order. Following each of the main open questions was a series of suggested probes which the interviewer was free to use or omit, depending on the fullness of the answer to the original question. The interview took between one and a half to two hours, on the average.

The interviews were recorded on an Audograph and later transcribed into typescript.

[1]This article is based on data drawn from a study on "Identification in Young Children," conducted by the Harvard Laboratory of Human Development under the direction of Professor Robert R. Sears, and is supported by a grant from the National Institute of Mental Health (M-461). The author wishes to thank Dr. Eleanor E. Maccoby for her help with the article.

On the basis of the number of records used to complete the standard interview, two extreme groups of 50 respondents each were chosen. The "long" interview group was made up of 40 people who talked for approximately two and a quarter hours, and 10 who talked for more than two and a half hours. The "short" interview group contained 25 respondents who completed the interview in less than one hour, and 25 respondents who took about one and a quarter hours to respond to all of the questions. The two extreme groups chosen for this comparison differed by approximately one hour, on the average, in the length of the interviews.[2]

The 100 interviews considered in this report were taken by ten different interviewers. Two interviewers were responsible for 29, or 58%, of the short interviews, and one other interviewer took 18, or 36%, of the long interviews. It appears, therefore, that the length of the interviews may have been a function of two interdependent factors: the interviewer's technique and the amount that the respondent talked.

Typescripts were made from the records, and 188 judgments, or codings, were made from each interview. Two raters independently made the judgments on each interview and then reached a "pooled" judgment by discussing their ratings. Raters were rotated so that each of a group of six judges worked with every other judge an equal number of times. For the various analyses discussed in the present study, five scales each of the I, S, and O types were selected.

Results

The comparison of short and long interviews is based on the assumption that even the short interviews contained sufficient information for the raters to make a judgment. Out of the total of 750 pooled judgments being considered for each group, only 14 judgments in each group were not codable.

Reliability of judgments. Since two coders made the judgments on each interview, the correlations between their ratings prior to pooling provide an index to the reliability of coding. In Table 1 it can be seen that the reliabilities on the item scales were in every instance higher for the short than for the long interviews. On two of the item scales the differences between the reliabilities are statistically significant, and on two other scales the differences approach significance.

Table 1

Two-Coder Reliability of Item Codes Based on Long and Short Interviews

	Scales				
	I-1	I-2	I-3	I-4	I-5
Long interviews	.50	.65	.65	.52	.50
Short interviews	.86	.82	.74	.76	.72
$t_{r_1 r_2}$	3.57**	1.83	.84	2.01*	1.72

** Significant at the 1% level.
* Significant at the 5% level.

From Tables 2 and 3 it can be seen that, with one exception, the reliabilities of judgments on summary and over-all codes are similar regardless of the length of response. Scale O-4 dealt with the "mother's esteem of father" and was one of the most highly inferential scales used. It is interesting to note that when the basis for the judgment was scattered throughout the interview, was not based on any specific questions, and involved a great deal of inference by the coder, long interviews, providing more material, permitted higher agreement among the judges. In general, though, when the code is based on a specific response, emendations and elaborations of the response detract from intercoder agreement. On codes encompassing more material in the interview, length of response has little influence on reliability.

Table 2

Two-Coder Reliability of Summary Codes Based on Long and Short Interviews

	Scales				
	S-1	S-2	S-3	S-4	S-5
Long interviews	.74	.60	.44	.77	.83
Short interviews	.66	.59	.59	.75	.74
$t_{r_1 r_2}$	.75	.07	.99	.23	1.14

Discriminability of judgments. In research interviews, on the basis of which coded variables will be related to other variables, it is a

[2]Since the recorder disc revolved at a constant rate, it was possible for a faster talking respondent to say more in a given length of time. Actually, the correlation between length of interview as determined by number of discs used and the number of pages of the typescript is .75. The choice of extreme groups should minimize this source of error.

prime requisite that the judgments discriminate maximally among respondents. With the variability of judgments as a criterion, the variances of the 15 codes on the long and short interview groups were computed, and the significance of difference between the "long" and "short" variances was computed. The results are given in Table 4. Of the 15 comparisons, three of the variances based on the long interviews are larger at the 5% level or better, and on one scale the variability of scores based on short interviews is significantly greater. However, the mean of the long interview variances is not significantly different from the mean variance of the short interview judgments (Festinger's $d = .86$) [2].

Table 3
Two-Coder Reliability of Over-all Codes Based on Long and Short Interviews

	Scales				
	O-1	O-2	O-3	O-4	O-5
Long interviews	.74	.58	.71	.72	.33
Short interviews	.82	.75	.75	.37	.42
$t_{r_1 r_2}$	.99	1.49	.41	2.49**	.50

** Significant at the 1% level.

Table 4
Variances of Judgments on Long and Short Interviews for Item (I), Summary (S), and Over-all (O) Codes

Scale	Variances long interviews	Variances short interviews	*F*
I-1	.89	1.14	1.28
I-2	1.41	1.58	1.12
I-3	1.01	.97	1.04
I-4	3.06	1.13	2.71**
I-5	1.00	1.16	1.16
S-1	.99	.94	1.05
S-2	2.40	2.53	1.05
S-3	.60	.70	1.17
S-4	1.33	.95	1.40*
S-5	2.25	1.78	1.26
O-1	1.70	1.56	1.09
O-2	.91	1.42	1.56*
O-3	2.15	2.20	1.02
O-4	1.87	1.44	1.30
O-5	7.82	5.07	1.54*

** Significant at the 1% level.
* Significant at the 5% level.

Validity of judgments. An ideal validity check would be to compare the judgments made on the interviews with the overt behaviors which the scales purport to measure. Since this was not feasible within the limits of the present study, a criterion of *logical* validity was used, that is, the utility of scores in generating predicted relationships with other variables. Based on the total 379 interviews, an analysis of social class differences in child-rearing practices was carried out.[3] It was found that 9 of the 15 scales being discussed in the current study were highly related to the social class memberships of the respondents. For example, there is more frequent use of physical punishment (Scale O-2) in the lower than in the middle class. Following the analysis of the total sample, specific predictions about the long and short interview groups were made as to how the variables would relate to social class membership. Within the limits of the reliabilities of the scales, it would be expected that the more valid the scales, the more clearly they will relate to social class.

To test these relationships, the long and short interviews were divided into higher and lower social class, and the means for each scale were calculated. In Table 5, the differences (d) between the means for the two social classes within each length-of-interview group are presented. The stronger the relationship to social class, the larger should be the d score. For example, for Scale I-4 (How Far is Child Expected to Go in School) the relationship is obviously higher for the short interviews, since there is a larger difference between the means, than in the long interviews (1.36 compared to .49).

If we sum the d scores separately for each type of scale, we find that there is a sizable d score for the item scales in favor of the short interview group; that is, the relationship of the variables to social class is higher for variables measured on the basis of the briefer interviews. These findings are in line with the higher reliabilities of the item scales, as reported in Table 1. It might be expected that, if the longer interviews permitted more valid judgments, the lower reliabilities of the long interview judgments might be somewhat overcome, thus equalizing the d scores. Such is obviously not the case.

[3] This study is reported in detail in [3].

Table 5

Means and Differences between Means on Selected Scales According to Length of Interview and Social Class of Respondent

Scale no.	Name	Short interviews: High social class ($N=25$)	Short interviews: Low social class ($N=25$)	Short interviews: d	Long interviews: High social class ($N=33$)	Long interviews: Low social class ($N=17$)	Long interviews: d
I-2	Permissiveness for masturbation	2.64	1.70	.94	2.74	1.91	.83
I-4	How far child expected to go in school	4.88	3.52	1.36	4.55	4.06	.49
I-5	Demonstrativeness: mother to child	3.38	3.32	.06	3.32	3.24	.08
				$\Sigma=2.36$			$\Sigma=1.40$
S-1	Severity of toilet training	2.36	2.32	.04	2.53	3.24	.71
S-2	Mother's sex anxiety	3.02	4.66	1.64	3.36	4.56	1.20
S-4	Permissiveness for aggression toward parents	2.40	1.64	.76	2.42	1.76	.66
				$\Sigma=2.44$			$\Sigma=2.57$
O-2	Use of physical punishment	3.16	3.90	.74	3.26	4.00	.74
O-4	Mother's evaluation of father	5.02	4.92	.10	5.17	4.26	.91
O-5	Agents of caretaking	3.26	2.20	1.06	3.89	3.94	.05
				$\Sigma=1.90$			$\Sigma=1.70$

On neither the summary nor over-all scales did added length increase the accuracy of prediction. Actually, three d scores favor the short interviews; two, the long interviews; and on one scale the d scores are equal. The sums of d scores on the S and O scales are very similar. Therefore, contrary to our original expectations, on none of the three types of scales did added length of response appear to increase the accuracy of predictions.

Summary and Conclusions

Fifty long and fifty short interviews conducted with a standardized open interview schedule were selected, and various characteristics of the coder judgments made from the interviews were compared. The short interviews yielded as many codable responses as the long interviews. The following findings emerged:

1. When judgments are made on a specific section of the interview, intercoder agreement is higher on the short interviews. There are no differences in reliabilities between judgments on long and short interviews when the appropriate material is based on a group of questions scattered throughout the interview.

2. There are no differences between the ability of long and the ability of short interview judgments to discriminate among respondents.

3. When validity is judged by the utility of the ratings in generating relationships with other variables, long and short interviews are equally valid on lengthy and scattered materials in the interviews. There is a suggestion that judgments on a specific section of the interview are more valid in the short interviews, although this may be a reflection of the higher reliabilities of these judgments.

Received December 4, 1953.

References

1. Feldman, J. J., Hyman, H., & Hart, C. W. Interviewer effects on the quality of survey data. *Publ. Opin. Quart.*, 1951-52, **15**, 734-761.
2. Festinger, L. The significance of difference between means without reference to the frequency distribution function. *Psychometrika*, 1946, **11**, 97-105.
3. Maccoby, Eleanor, & Gibbs, P. Child rearing and social class. In C. B. Stendler & W. Martin (Eds.), *Readings in child development.* New York: Harcourt, Brace, 1954. Pp. 380-396.
4. Shapiro, J., & Eberhart, J. C. Interviewer differences in an intensive interview survey. *Int. J. Opin.-Attitude Res.*, 1947, **1**, 1-17.

NEW BOOKS AND TESTS

Books

Carmichael, Leonard. (Ed.) *Manual of child psychology.* (2nd Ed.) New York: Wiley, 1954. Pp. ix + 1295. $12.00.

The increments of knowledge from eight more years of research fully justify the publication of a new edition of the *Manual of Child Psychology.* The authors of the chapters differ considerably, however, in the extent to which they perceive forward movements in their fields. Some chapters are rewritten extensively with the citation of many new studies and the reappraisal of some older ones; others seem to have undergone only perfunctory changes since the 1946 edition. Among the most thoroughly revised sections are Munn's account of learning, McCarthy's of language development, Harold Jones's of the environment and development, Vernon Jones's of character, and Jersild's of emotion. Chapters almost unchanged are those by Carmichael on prenatal development, Cruikshank on animal infancy, and Pratt on the neonate. One wonders whether the statement in the preface that "development has been more active in certain fields" accounts for all of the differences. Three chapters are new, one of them an excellent account of adolescence by Horrocks. The other new chapters do not fulfill their potentialities. Benda, writing on the psychopathology of childhood, devotes 31 pages to a well-documented survey of mental deficiency, and then gives only 14 pages to a didactic and elementary account of psychoneuroses and childhood schizophrenia, citing hardly a research study. Harold and Gladys Anderson append over 200 references to their chapter on social development but use them mainly in lists of names and dates. The chapter is expository with a few illustrations from the authors' own research. The *Manual* as a whole, in spite of the inevitable unevenness resulting from multiple authorship, remains a useful guide to the research on childhood. Its direct value to the needs and concerns of clinical psychologists is limited. — *L. F. S.*

Note.—The reviews were prepared by the Editor and Associate Editors, who may be identified by their initials.

Gladwin, Thomas, & Sarason, Seymour B. *Truk: man in paradise.* New York: Wenner-Gren Foundation for Anthropological Research, 1953. Pp. 651 + 8 plates. $6.50.

Long aware of the values of anthropological studies for their work, clinicians will find this monograph particularly relevant and stimulating. In the first place, it represents at a still more sophisticated and insightful level the kind of collaboration between anthropologist and psychologist that made the Dubois-Oberholzer study of the Alorese so seminal for both disciplines. Second, since it is one of a series of investigations of Trukese culture, this report is unburdened with such specifically ethnological matters as linguistics, formal social structure, systems of property ownership, and the character of the material culture. Its content is concerned entirely with Trukese personality development and the sociocultural matrix within which it finds expression. Consequently, it deals with the problems of greatest moment to American clinical psychologists but from a point of view that permits a valuable perspective. Third, the book is crammed with extensive life histories and excerpts from projective test protocols. These materials will excite the imagination of all who use such data in their daily work and provoke fruitful hypotheses among the thoughtful for testing in both clinical and research contexts. This fat volume is at once lively and serious, and it contributes both to an understanding of the conditions affecting personality development and of the techniques by which it is conceptualized and investigated. — *E. J. S.*

Greenacre, Phyllis. (Ed.) *Affective disorders.* New York: International Universities Press, 1953. Pp. 212. $3.00.

A symposium of five papers, mainly on depression, presented before the American Psychoanalytic Association in 1950 and 1951. Edward Bibring and Edith Jacobson are in agreement that depression is a helplessness reaction with loss of self-esteem, and both have difficulty in reconciling the view with some earlier analytic interpretations. Elizabeth Zetzel presents an extended apology for Melanie Klein's theories. George Gero gives an interesting case of anorexia as an equivalent of depression. M. Katan views both depressed and man-

ic symptoms as mechanisms of denial and restitution. — *L. F. S.*

Harms, Ernest. *Essentials of abnormal child psychology.* New York: Julian Press, 1953. Pp. xiii + 265. $5.00.

In this book, the author presents his views of the genetic development of the child, before discussing therapy and diagnosis. He considers the concept of infantile sexuality quite inadequate, and the parental relation to be beyond sexual experience. The relationship to the father is described under the rubric of *paternus* and to the mother of *materna,* which he considers to be inherited archetypical factors comparable to Jung's *animus* and *anima.* His discussion of case material and specific diagnostic categories follows his conviction that different techniques are applicable to different cases, and that patients are best helped by psychotherapists who share their own psychic structure. He believes the incidence of true childhood schizophrenia greatly overestimated. His discussion of the diagnosis and treatment of feeblemindedness, and of the use of child art as diagnostic means, will be of particular interest to psychologists. — *A. R.*

Pennington, L. A., & Berg, Irwin A. (Eds.) *An introduction to clinical psychology.* (2nd Ed.) New York: Ronald, 1954. Pp. vii + 709. $6.50.

This is an excellent textbook, planned to give students an overview of the whole field of clinical psychology. Its editors have emphasized general principles rather than the mastery of special techniques. Although there is a long list of contributing writers, their varying viewpoints do not distract from the basic philosophy but rather make the student aware of the complexity of the field of clinical psychology. The second edition is not a reprint but a carefully considered revision. The chapters that are retained have been revised, and at least ten new chapters are added. It should be valuable for an introductory course in clinical psychology. — *B. M. L.*

Pepinsky, Harold B., & Pepinsky, Pauline Nichols. *Counseling theory and practice.* New York: Ronald, 1954. Pp. viii + 307. $4.50.

The Pepinskys resolve a familiar dilemma by integration. Counselors can and must be scientists as well as practitioners; they do not have to choose between the roles. Furthermore, being a scientist is not limited to an isolated part of one's life labeled "research," but is an essential feature of the counseling process itself. Like an experiment, each client presents a challenge to observe, to infer, to relate to a general framework of theory, to formulate particular hypotheses, and to assess the outcome. The authors wisely refrain from preaching an acceptance of their analysis. Instead, they illustrate it persuasively. The place of formal theory is shown by a learning approach to client behavior which uses Hull's symbols and also draws heavily on Mowrer, Berlyne, Skinner, Dollard and Miller, and Brown and Farber. Observation, inference, hypothesis making, and assessment are illustrated by case studies of counseling. Assessment is approached realistically, with a cogent discussion of criterion variables. This book may help to make counselors—and psychotherapists—more aware of the assumptions that they bring to their contacts with clients. As a text, it may help students to see the intimate relationships between science and practice. — *L.F.S.*

Stott, D. H. *Saving children from delinquency.* New York: Philosophical Library, 1953. Pp. x + 266. $4.75.

Stott has placed the emphasis upon "saving" children from delinquency—by prevention if possible, and by avoiding further difficulty after delinquency has occurred. The illustrations and situations are typically British, but the American reader will find much to help him in meeting his own problems since the basic needs are similar. Visiting teachers, probation officers, and social workers will find the book very readable and stimulating. — *B. M. L.*

Vedder, Clyde B. *The juvenile offender: perspective and readings.* Garden City, N. Y.: Doubleday, 1954. Pp. xii + 510. $6.00.

As the subtitle indicates, this book is a collection of readings from periodicals dealing with the subject of juvenile delinquency. The 71 reprinted articles are organized into 13 chapters, under headings like "The Extent of Juvenile Delinquency," "Economic Conditions and Familial Factors," "Special Personality and Behavior Problems," "Probation," and the like. Each chapter starts with a brief discussion of the particular aspect of juvenile delinquency by the author of this volume, well documented from authoritative treatments, and followed by the periodical readings. Twelve social science periodicals are represented in the reprints of the 71 articles, but 90% are from four journals, and 60% from one: *Federal Probation.* In reading the volume, the impression is gained that the entire book is a republication of the last-mentioned periodical. Nevertheless, the variety of articles is great, and most of the authors are outstanding authorities in the field. It is a valuable source book for the student of juvenile delinquency. — *M. K.*

Wolffheim, Nelly. *Psychology in the nursery school.* (Trans. by Charles L. Hannam.) New York: Philosophical Library, 1953. Pp. 144. $3.75.

Psychology in the Nursery School is a translation of a work originally published in German which pleads for a psychoanalytically oriented nursery school. Dr. Wolffheim feels that Freudian psychology has an important contribution to the education of young children, and should lead to new ways of studying them. She believes that psychoanalysis would benefit both the teacher and her educational work; in fact she feels it would be a de-

sirable part of the professional training of teachers. The book is challenging but not very practical. The author admits that her point of view is based on theoretical studies and experiences in small nursery schools only. — *B. M. L.*

Tests

Burgemeister, Bessie B., Blum, Lucille Hollander, & Lorge, Irving. *Columbia Mental Maturity Scale.* Ages 3-12. Individual test. 1 form. Untimed, (15-30) min. Kit of 100 test cards, with manual, pp. 13 ($35.00); record blank (85¢ per 35). Yonkers, N. Y.: World Book Co., 1954.

A seven-year research project culminated in the development of this unique individual intelligence test designed for the needs of children with cerebral palsy and other severe motor or speech handicaps. The materials are 100 large cards, 6 by 19 in., each containing from three to five drawings, many in color. The task is the same for each item: to select the picture that does not belong with the others. Most children point to the figure; handicapped children who can only nod or make a sound can also be examined successfully. The test was standardized on 957 normal children from 3 through 12 years of age, and also tried with 147 cerebral palsied children. In single-year age groups of normal children, its corrected odd-even reliability is from .89 to .92, and its correlations with the Stanford-Binet range from .66 to .88, with an average value of about .75. The test will surely find wide use in clinical work and research with the severely handicapped. It is not likely to replace existing instruments for testing normal children. The very features of simplicity that make it suitable for the handicapped hinder the useful auxiliary observations often provided by verbal tests and performance tests with varied tasks. Even so, another well-standardized intelligence test is a real asset to clinical psychology, and may find uses beyond its authors' original intentions. — *L. F. S.*

Eisenson, John. *Examining for aphasia.* (Rev. Ed.) Adolescent-adult. Individual test. 1 form. Manual with plates, pp. v + 73 ($3.50); record form, pp. 12, ($3.50 per 25); set ($6.00). New York: Psychological Corp., 1954.

Eisenson's revised tests for aphasia do not differ greatly from their first edition of 1946. The broad scheme of testing remains the same, with some improvements in the pictorial materials, in the arrangement of plates, and in the selection of paragraphs for oral and silent reading. An examination for aphasia is necessarily qualitative, and the tests therefore cannot be judged by the usual standards for reliability and validity. Some approximate quantifications of these tests would be helpful, however, at least to give an examiner clues for judging whether the disability in a given area is "moderate" or "severe." The manual is more extensively revised than the test itself, the first 28 pages comprising a miniature textbook on nature, symptoms, and varieties of aphasia, and the examination of aphasic children. Perhaps because of its condensation, this introductory material tends to be arbitrary and uncritical. Still, it provides a useful check list of definitions and procedures for the use of experienced psychologists. — *L. F. S.*

Wesman, Alexander G., *Verbal Test,* & Doppelt, Jerome E., *Numerical Test. Personnel Test for Industry (PTI).* Trade school-adult. 2 forms, each test. 5 (10) min., verbal; 20 (25) min., numerical. Test booklet, each test ($1.60 per 25), with manual, pp. 10, and keys; specimen set (35¢). New York: Psychological Corp., 1952, 1954.

The verbal and numerical tests introduce a new series of *Personnel Tests for Industry* which is to include tests of other abilities and skills. The verbal test is a 50-item, multiple-choice vocabulary test, and the numerical test a series of 30 arithmetic problems based on industrial tasks. The manual, modestly designated as preliminary, gives extensive information about reliabilities and the prediction of criteria, based on industrial workers, and percentile norms derived from several industrial samples and one trade school sample, totaling some 1400 cases. Although conventionally conceived, the tests show the application of high professional standards. They illustrate the trend toward the construction of tests especially designed for adults in industry. — *L. F. S.*

Books Received

Abramson, Harold A. (Ed.) *Problems of consciousness.* New York: Josiah Macy, Jr. Foundation, 1954. Pp. 177. $3.25.

Anastasi, Anne. *Psychological testing.* New York: Macmillan, 1954. Pp. xiii + 682. $6.75.

Burmeister, Eva. *Roofs for the family.* New York: Columbia Univer. Press, 1954. Pp. 203. $3.25.

Caplow, Theodore. *The sociology of work.* Minneapolis: Univer. of Minnesota Press, 1954. Pp. viii + 330. $5.00.

Commins, W. D., & Fagin, Barry. *Principles of educational psychology.* New York: Ronald, 1954. Pp. xvi + 795. $5.75.

De Sade. *Selected writings.* (Trans. by Leonard de Saint-Yves.) London: Peter Owen, and New York: British Book Centre, 1954. Pp. 306. $6.75.

Freud, Sigmund. *The origins of psychoanalysis: letters to Wilhelm Fliess, drafts and notes: 1887-1902.* New York: Basic Books, 1954. Pp. xi + 486. $6.75.

Fryer, Douglas H., Henry, Edwin R., & Sparks, Charles P. *General psychology.* (4th Ed.) College Outline Series. New York: Barnes & Noble, 1954. Pp. xix + 300. $1.50 (paper).

Gorer, Geoffrey. *The life and ideas of the Marquis de Sade.* London: Peter Owen, and New York: British Book Centre, 1954. Pp. 244. $3.50.

Martin, John Bartlow. ***Break down the walls.*** New York: Ballantine, 1954. Pp. 310. $3.50 (cloth), 50¢ (paper).

Pitt, William J., & Goldberg, Jacob A. ***Psychology.*** New York: McGraw-Hill, 1954. Pp. 414. $4.50.

Remmers, H. H., Ryden, Einar R., & Morgan, Clellen L. ***Introduction to educational psychology.*** New York: Harper, 1954. Pp. ix + 435. $4.00.

Runes, Dagobert D. ***Letters to my daughter.*** New York: Philosophical Library, 1954. Pp. 131. $2.50.

Ryan, Thomas Arthur, & Smith, Patricia Cain. ***Principles of industrial psychology.*** New York: Ronald, 1954. Pp. xiv + 534. $5.50.

Schwing, Gertrud. ***A way to the soul of the mentally ill.*** New York: International Universities Press, 1954. Pp. 158. $3.00.

Wulfeck, Joseph W., & Bennett, Edward M. ***The language of dynamic psychology as related to motivation research.*** New York: McGraw-Hill, 1954. Pp. 111. $4.00.

JOURNAL OF CONSULTING PSYCHOLOGY
AVAILABLE BACK ISSUES

July 1, 1954

The issues listed below are available for sale

YEAR	VOLUME	FEB	APR	JUN	AUG	OCT	DEC	PRICE PER NUMBER	PRICE PER VOLUME
1937	1	1	2	3	4	5	6	$1.50	$8.00
1938	2	–	2	3	4	5	6	$1.50	——
1939	3	1	2	3	4	5	6	$1.50	$8.00
1940	4	1	2	3	4	5	–	$1.50	——
1941	5	1	2	3	4	5	6	$1.50	$8.00
1942	6	1	–	3	–	–	6	$1.50	——
1943	7	–	2	3	4	5	6	$1.50	——
1944	8	1	–	3	4	5	6	$1.50	——
1945	9	–	–	–	4	5	6	$1.50	——
1946	10	1	2	3	4	5	–	$1.50	——
1947	11	–	–	–	4	5	6	$1.50	——
1948	12	1	2	3	4	5	6	$1.50	$8.00
1949	13	1	2	3	4	5	6	$1.50	$8.00
1950	14	1	2	3	4	5	6	$1.50	$8.00
1951	15	1	2	3	4	5	6	$1.50	$8.00
1952	16	1	2	3	4	5	6	$1.50	$8.00
1953	17	1	2	3	4	5	6	$1.50	$8.00
1954	18	By subscription $7.00, foreign $7.50						$1.50	——

Postage prepaid on U. S. orders. Add $.50 per volume on foreign orders. All stock subject to prior sale.

The American Psychological Association offers the following discounts on orders for any one journal:

10% on orders of $50.00 and over
20% on orders of $100.00 and over
30% on orders of $150.00 and over

Address subscriptions and orders for back issues to:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street N. W.
Washington 6, D. C.

PSYCHIATRY

JOURNAL FOR THE STUDY OF INTERPERSONAL PROCESSES

Contents of No. 2, Vol. XVII, May 1954

MABEL BLAKE COHEN, GRACE BAKER, ROBERT A. COHEN, FRIEDA FROMM-REICHMANN, AND EDITH V. WEIGERT: *An Intensive Study of Twelve Cases of Manic-Depressive Psychosis.*

JULES HENRY: *The Formal Social Structure of a Psychiatric Hospital.*

OSCAR LEGAULT: *Denial as a Complex Process in Post Lobotomy.*

ROBERT N. WILSON: *Poetic Creativity.*

SHIRLEY W. KAHN AND A. RODNEY PRESTWOOD: *Group Therapy of Parents as an Adjunct to the Treatment of Schizophrenic Patients.*

ELINOR W. DEMAREST AND ARTHUR TEICHER: *Transference in Group Therapy.*

EDITORIAL NOTES.

BRIEF COMMUNICATIONS.

BOOK REVIEWS.

Published in February, May, August, and November by the William Alanson White Psychiatric Foundation, 1711 Rhode Island Ave., N. W., Washington 6, D. C.

Subscription price $10.00 for 1954; foreign $10.80
Single copy $2.50; foreign $2.70

The Foundation announces the publication by W. W. Norton & Company of *The Psychiatric Interview* by Harry Stack Sullivan. This book discusses and analyzes the technique of carrying on a personal interview and provides the first complete statement of how Sullivan applied to the practical work of interviewing the conceptual framework set forth in his earlier books. Copies may be purchased from the William Alanson White Psychiatric Foundation, 1711 Rhode Island Ave., N. W., Washington 6, D. C. $4.50 prepaid.